Maleness-on-the-Y (MoY) orchestrates male sex determination in major agricultural fruit fly pests

Meccariello, Angela ; Salvemini, Marco ; Primo, Pasquale ; Hall, Brantley ; Koskinioti, Panagiota ; Dalíková, Martina ; Gravina, Andrea ; Gucciardino, Michela Anna ; Forlenza, Federica ; Gregoriou, Maria-Eleni ; Ippolito, Domenica ; Monti, Simona Maria ; Petrella, Valeria ; Perrotta, Maryanna Martina ; Schmeing, Stephan ; Ruggiero, Alessia ; Scolari, Francesca ; Giordano, Ennio ; Tsoumani, Konstantina T ; Marec, František ; Windbichler, Nikolai ; Arunkumar, Kallare P ; Bourtzis, Kostas ; Mathiopoulos, Kostas D ; Ragoussis, Jiannis ; Vitagliano, Luigi ; Tu, Zhijian ; Papathanos, Philippos Aris ; Robinson, Mark D ; Saccone, Giuseppe

Abstract: In insects, rapidly evolving primary sex-determining signals are transduced by a conserved regulatory module controlling sexual differentiation. In the agricultural pest Ceratitis capitata (Mediterranean fruit fly, or Medfly), we identified a Y-linked gene, Maleness-on-the-Y (MoY), encoding a small protein that is necessary and sufficient for male development. Silencing or disruption of MoY in XY embryos causes feminization, whereas overexpression of MoY in XX embryos induces masculinization. Crosses between transformed XY females and XX males give rise to males and females, indicating that a Y chromosome can be transmitted by XY females. MoY is Y-linked and functionally conserved in other species of the Tephritidae family, highlighting its potential to serve as a tool for developing more effective control strategies against these major agricultural insect pests.

DOI: https://doi.org/10.1126/science.aax1318

Maleness-on-the-Y (*MoY*) orchestrates male sex determination in major agricultural fruit fly pests

Angela Meccariello[1,†], Marco Salvemini[1,†], Pasquale Primo[1], Brantley Hall[2], Panagiota Koskinioti[3,4], Martina Dalíková[5,6], Andrea Gravina[1], Michela Anna Gucciardino[1], Federica Forlenza[1], Maria-Eleni Gregoriou[4], Domenica Ippolito[1], Simona Maria Monti[7], Valeria Petrella[1], Maryanna Martina Perrotta[1], Stephan Schmeing[8], Alessia Ruggiero[7] Francesca Scolari[9], Ennio Giordano[1], Konstantina T. Tsoumani[4], František Marec[5], Nikolai Windbichler[10], Kallare P. Arunkumar[11], Kostas Bourtzis[3], Kostas D. Mathiopoulos[4], Jiannis Ragoussis[12], Luigi Vitagliano[7], Zhijian Tu[2], Philippos Aris Papathanos*[13,14], Mark D. Robinson*[8] and Giuseppe Saccone[1]*

[1] Department of Biology, University of Naples "Federico II", 80126, Napoli, Italy

[2] Department of Biochemistry, Virginia Tech, Blacksburg, United States.

[3] Insect Pest Control Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, A-1400 Vienna, Austria.

[4]Department of Biochemistry and Biotechnology, University of Thessaly, Larissa, Greece.

[5]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, 370 05 České Budějovice, Czech Republic.

[6]Faculty of Science, University of South Bohemia, 370 05 České Budějovice, Czech Republic.

[7]Institute of Biostructures and Bioimaging (IBB), CNR, 80134 Naples, Italy.

[8]Institute of Molecular Life Sciences and SIB Swiss Institute of Bioinformatics, University of Zurich, 8057 Zurich, Switzerland.

[9]Department of Biology and Biotechnology, University of Pavia, Pavia, 27100, Italy.

[10] Department of Life Sciences, Imperial College London, South Kensington Campus, SW7 2AZ, United Kingdom.

[11] Centre of Excellence for Genetics and Genomics of Silk moths, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India.

[12] McGill University and Genome Quebec Innovation Centre, Department of Human Genetics and Bioengineering, Montreal H3A 0G1, Canada.

[13] Section of Genomics and Genetics, Department of Experimental Medicine, University of Perugia, 06132, Italy.

[14] Department of Entomology, The Robert H Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, 76100, Israel.

*Corresponding authors: giuseppe.saccone@unina.it, mark.robinson@imls.uzh.ch, p.papathanos@mail.huji.ac.il.

[†]These authors contributed equally to this work.

Abstract

In insects, rapidly evolving primary sex-determining signals are transduced by a conserved regulatory module controlling sexual differentiation. In the agricultural pest *Ceratitis capitata* we identified a Y-linked gene, *Maleness-on the-Y* (*MoY*), encoding a small protein that is necessary and sufficient for male development. Silencing or disruption of *MoY* in XY embryos causes feminization whereas overexpression of *MoY* in XX embryos induces masculinization. Notably, crosses between transformed XY females and XX males give rise to males and females, indicating that a Y chromosome can be transmitted by XY females. Interestingly, *MoY* is Y-linked and functionally conserved in other species of the Tephritidae family, highlighting its potential to serve as a tool for developing more effective control strategies against these major agricultural insect pests.

One Sentence Summary

The Medfly male determining factor is the Y-linked gene *MoY* encoding a small, novel protein that is functionally conserved in major fruit fly pests.

Tephritidae is a dipteran family comprising 5000 species, dozens of which are invasive and highly relevant pests of fruit crops. *Ceratitis capitata* (Medfly) is one of the most destructive members of this taxon, affecting over 200 plant species (*1*). Besides pesticides, the most successful method to control Medfly is the Sterile Insect Technique (SIT) (*2*), which involves the continuous mass-release of biofactory-reared, sterilized males that suppress wild populations by mating with wild females. A key determinant to the success of SIT programs has been the translocation of selectable traits to Medfly Y chromosome in genetic sexing strains that enable male selection on a massive scale (*2*). However, the development of similar strains in other Tephritidae pest species has been difficult using classical genetics, inhibiting the application of SIT. Identifying the *M* factor in Medfly and in related pests holds significant promise for the development of novel genetic sexing strains using modern genetics (*2*), or even for transforming females into males thereby increasing the efficiency of insect bio-factories.

In insects, widely divergent primary signals of sex determination act *via* the conserved genetic switch *transformer* (*tra*), which was first characterized in *Drosophila*, as a gene regulated by and operating through sex-specific alternative splicing (*3–6*). In females, two doses of the X

chromosome result in an early zygotic transcriptional burst of the master gene *Sex-lethal* (*Sxl*) which promotes female-specific splicing of *tra* and female differentiation. In males, a single X leads by default to a *tra* transcript that encode a short non-functional TRA polypeptide and male differentiation occurs. In Medfly, as in other non-Drosophilidae species, *Sxl* is not involved in sex determination (*3*). Unlike *Drosophila*, maternal deposition of *Ceratitis capitata tra* (*Cctra*) in developing embryos initiates its positive autoregulatory female-specific splicing, similarly to the housefly (*5*, *6*), leading to female differentiation. In Medfly XY embryos, a Y-chromosome linked *M* factor either directly or indirectly represses *Cctra* function thus promoting male development (*5*). However, the molecular identity of this Y-linked *M* factor has remained unknown (*7*, *8*).

We conducted the search for this *M* factor by: 1) focusing on transcripts in 4-8h hours (h) after egg laying (AEL), the period when male sex determination is first established (*9*); 2) producing an XX-only embryonic RNAseq dataset as a reference; 3) developing a long read PacBio-sequencing-based male genome assembly from the *Fam18* Medfly strain (*8*) that bears a shorter Y chromosome; and 4) searching for conservation of putative *M* factors in another Tephritidae, the olive fruit fly *Bactrocera oleae*. We identified 19 *M* factor candidates expressing transcripts in mixed-sex embryos but not in XX-only embryos (Fig. 1A; table S1; supplementary text S1). Seven of these transcripts did not map to the *Fam18* male genome assembly and were excluded from further analysis. Sequence similarity searches by BLASTn showed that three out of the remaining 12 transcripts had hits to XY but not to XX embryonic transcripts from *B. oleae*. Furthermore, one of these three Medfly transcripts (DN40292_c0_g3_i1) corresponds to a 0.7kb sequence we previously identified in a preliminary screen by the chromosome quotient approach (CQ), which predicts Y-linkage of a sequence based on the ratio of genomic reads mapping from female *versus* male samples (*10*) (supplementary text S2). This transcript mapped to a predicted 12kb long Y-linked contig in the *Fam18* genome assembly. Functional analysis (below) confirmed that this gene is the Medfly *M* factor and was thus named *Maleness-on-the-Y* (*MoY*). *MoY* is located on the long arm of the Y chromosome in proximity to the centromere (Fig. 1B) in a genomic region that contains 9 other transcription units (Fig. 1C; supplementary text S3). *MoY* expression begins at 2-3h AEL, prior to embryonic cellularization, peaks at 15h and becomes undetectable by 48h until adulthood (Fig. 1D; Data S1).

Transient silencing of *MoY* by injecting double-stranded RNA into 0-1h AEL embryos (embryonic RNA interference, eRNAi) resulted in the loss of male-specific *Cctra* transcripts in 8h AEL

embryos (Fig. 2A-C; tables S2 and S3). We observed a switch to the female-specific *Cctra* splicing in 3-day-old XY larvae from injected embryos and in XY adult intersexes (fig. S1). Among adults, 38% (14/37) of the molecularly karyotyped XY individuals displayed complete phenotypic feminization, and 19% (7/37) were intersexes (Fig. 2D; fig. S2). We assigned phenotypic sex based on the presence or absence of the pair of orbital bristles on male heads, of the male genital apparatus and of the female-specific ovipositor (Fig. 2D). Individuals were phenotypically classified as intersexes if they displayed either a female-specific ovipositor together with male-specific orbital bristles, or male genital apparatus with a female head lacking the male-specific bristles. To evaluate the fertility of XY females and to test whether a Y chromosome can be maternally transmitted, crosses were established to XX males that were generated by eRNAi targeting *Cctra* (*5*) (table S4, fig. S3A). These crosses simplified the tracking of Y chromosomes excluding their paternal transmission. Among the 14 recovered XY females, one was fertile and transmitted the Y chromosome to one son. We found another XY female that transmitted the Y chromosome in a second experiment of *MoY*-eRNAi, using a strain displaying sex-specific pupal colours (*Vienna8*) (*2*) (table S2). To our knowledge, this is the first genetic demonstration that the two sexes of an animal species defined by each karyotype can be reciprocally switched in both directions while maintaining fertility. This highlights a surprising resilience of Medfly somatic and gonadal development in both sexes to perturbations in sex determination signaling and suggests that the Y chromosome has no major detrimental effects on the development and fertility of XY females.

To further evaluate the role of *MoY* in sex determination, loss-of-function alleles were generated using Cas9 ribonucleoproteins (*11*) targeting the MOY coding sequence (table S2; fig. S4A). Indels near the single guide RNA (sgRNA) target site were induced in the genomes of 4 G_0 XY larvae and 3 G_0 XY adult intersexes (fig. S5A). 50% (7/14) of the XY individuals (table S2) were transformed either into phenotypic females (2/14) or intersexes (5/14) (fig. S4B-D and fig. S5B). One XY female crossed to XX males was fertile and produced female-only G_1 offspring composed of 3 XY and 18 XX flies (table S5; fig. S3B and fig. S4D-E). Two of these XY G_1 females were analysed and both showed *MoY* frameshift-inducing deletions resulting in truncated MOY proteins (fig. S5A).

Next, we investigated whether *MoY* is sufficient for male sex determination. A 5kb genomic fragment, encompassing the *MoY* locus and flanking regulatory regions (Fig. 1C), was injected as

a linear PCR product or as circular plasmid into embryos (table S2). Male-specific *Cctra* splicing was induced in XX individuals at embryonic, larval and adult stages (Fig. 2C and fig. S6) and led to partial or full masculinization of up to 75% of XX flies (9/12) (table S2; Fig. 2D, fig. S6 and fig. S7). Similarly, microinjection of MOY recombinant protein into XX-only embryos led to partially masculinized flies (showing either male-specific orbital bristles or male genitalia) in 19% (6/31) of the emerged XX adults (table S2; fig. S8).

Ceratitis MoY DNA and protein sequences showed no significant BLAST hits to NCBI databases, suggesting novelty or high sequence divergence. In contrast, tBLASTn searches of available genomic or transcriptomic datasets from 14 Tephritidae species spanning 111 million years of evolution (*12*) identified putative MOY orthologues in 8 of them (Fig. 3A) with an average 41-57% amino acid sequence similarity to *Ceratitis* (supplementary text S4-S12). The most conserved portion is located in the N-terminal region where a consensus hexapeptide KXNSRT occurs (Fig. 3B). We confirmed by PCR of male and female genomic DNA the Y-linkage of *MoY* orthologues in the 4 of the 8 species, namely *B. oleae*, *B dorsalis*, *B. tryoni* and *B. jarvisi*, whose fly samples were available (Fig. 4A). Importantly, *MoY*-eRNAi orthologues in *B. oleae* and *B. dorsalis* led to feminization of XY flies (57% and 33% respectively), confirming the functional conservation of *MoY* (table S2; Fig. 4B and fig. S9).

Here, we demonstrate that *MoY* is the Y-linked *M* factor in *C. capitata* because it is both necessary and sufficient to initiate male development during embryogenesis. How *MoY* suppresses female-specific *Cctra* splicing and whether this regulation is direct or indirect remain unclear. MOY does not bear any similarity to known proteins, in contrast to *M* factors of *Aedes* and *Musca,* which are related to splicing factors (*13*, *14*). This finding suggests that *MoY* is a newly-emerged *M* factor. Surprisingly, we also discovered that *MoY* is functionally conserved in various Tephritidae species, in contrast to previous studies conducted on other insects whose *M* factors diverged rapidly, even within the same species (*13–16*). Our finding that masculinization through transient *MoY* misexpression in Medfly XX embryos does not lead to lethality as in other species (*15*, *16*) suggests that *MoY* is not involved in dosage compensation. However, it is not clear from the present data whether Medfly has such regulatory mechanisms to equalize expression of X-chromosome genes in both sexes. From a translational perspective, these features, including *MoY* conservation and its ability to fully masculinize XX individuals, make *MoY* a promising candidate for

transferring to other important Tephritidae pests established genetic control strategies such as SIT, and for future development of emerging methods such as male-converting gene drive (*17–20*).

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Javaregowda Nagaraju with whom the project was initiated. We thank Lino C. Polito, who initiated the Medfly research in Naples and led the first studies on Medfly sex determination and L. Gaudio for very stimulating encouragements. We thank Carlos Caceres, Sohel Ahmad, Jennifer Morrow for fly samples and methods, Bernardo Carvalho and Serena Aceto for bioinformatic support, Daniel Bopp, Alfredo Pagliuca and Lenny Rabinow for discussions and manuscript reviewing. This study benefited from discussion during Coordination Research Projects (CRP) supported by Joint FAO/IAEA Division. This research was supported by grants R-Dept-Biol-2015-18 to GS and MS, STAR2013_25 to MS, fellowships to AM and PP from the University of Naples Federico II. MDR acknowledges support from the SNSF (project numbers 31003A_143883, 31003A_175841) and SS and MDR from the University of Zurich Research Priority Program Evolution in Action. NW was funded by the BBSRC under the research grant BB/P000843/1. ZT was supported by NIH grants AI121284 and AI123338. PAP was supported by RLM award from MIUR (D.M. no. 79 04.02.2014). JR was supported by Compute Canada Resource Allocation Project (WST-164-AB), and Genome Canada Genome Innovation Node (244819) grants. KTT was supported by a Hellenic IKY Fellowship, and M-EG and KDM by OMIC-Engine, a Greek Research Infrastructure Program on Synthetic Biology. MD and FM acknowledge support from the Czech Science Foundation (grants 17-17211S and 17-13713S, respectively). Author contributions: see Supplementary materials.

SUPPLEMENTARY MATERIALS

Author contributions

Data and materials availability

Materials and Methods

Supplementary text S1 to S12

Figs. S1 to S9

Tables S1 to S9

Data S1 to S4

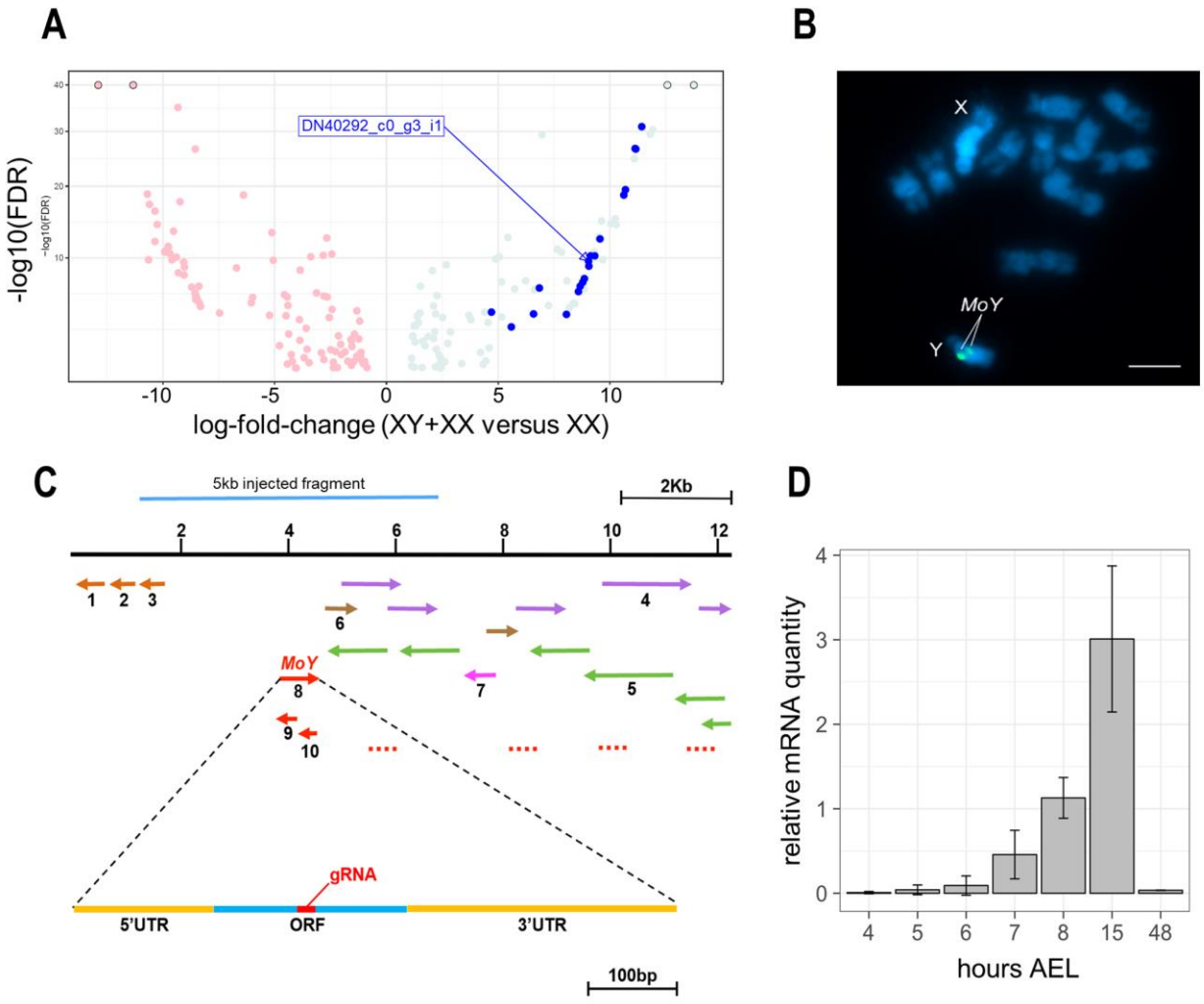

Fig. 1. *MoY* gene is Y-specific and transiently transcribed in embryos. (**A**) Volcano plot of 195 differentially expressed transcripts at 4-8 h AEL. Pink dots indicate XX-biased transcripts, light-blue dots indicate XX/XY-biased transcripts. Dark-blue dots indicated 19 transcripts with CQ value of 0. The transcript corresponding to *MoY* (DN40292_c0_g3_i1) is also shown. (**B**) Fluorescence *in situ* hybridization of *MoY* (green signals) on mitotic chromosomes stained with DAPI (blue); signals (one for each sister chromatid) locate *MoY* on the long arm of the Y chromosome near the centromere. (**C**) A scheme of a 12kb Y-linked genomic contig (contig00013010) containing the *MoY* transcript and other flanking transcriptional units (see supplementary text S3 for further details). Also shown is the 5kb region used for injections and the *MoY* guide RNA target site for Cas9. (**D**) Relative transcript expression of *MoY* during Medfly

embryogenesis compared to a housekeeping genes (AEL). For reference, Medfly cellularization occurs around 9 hours AEL.

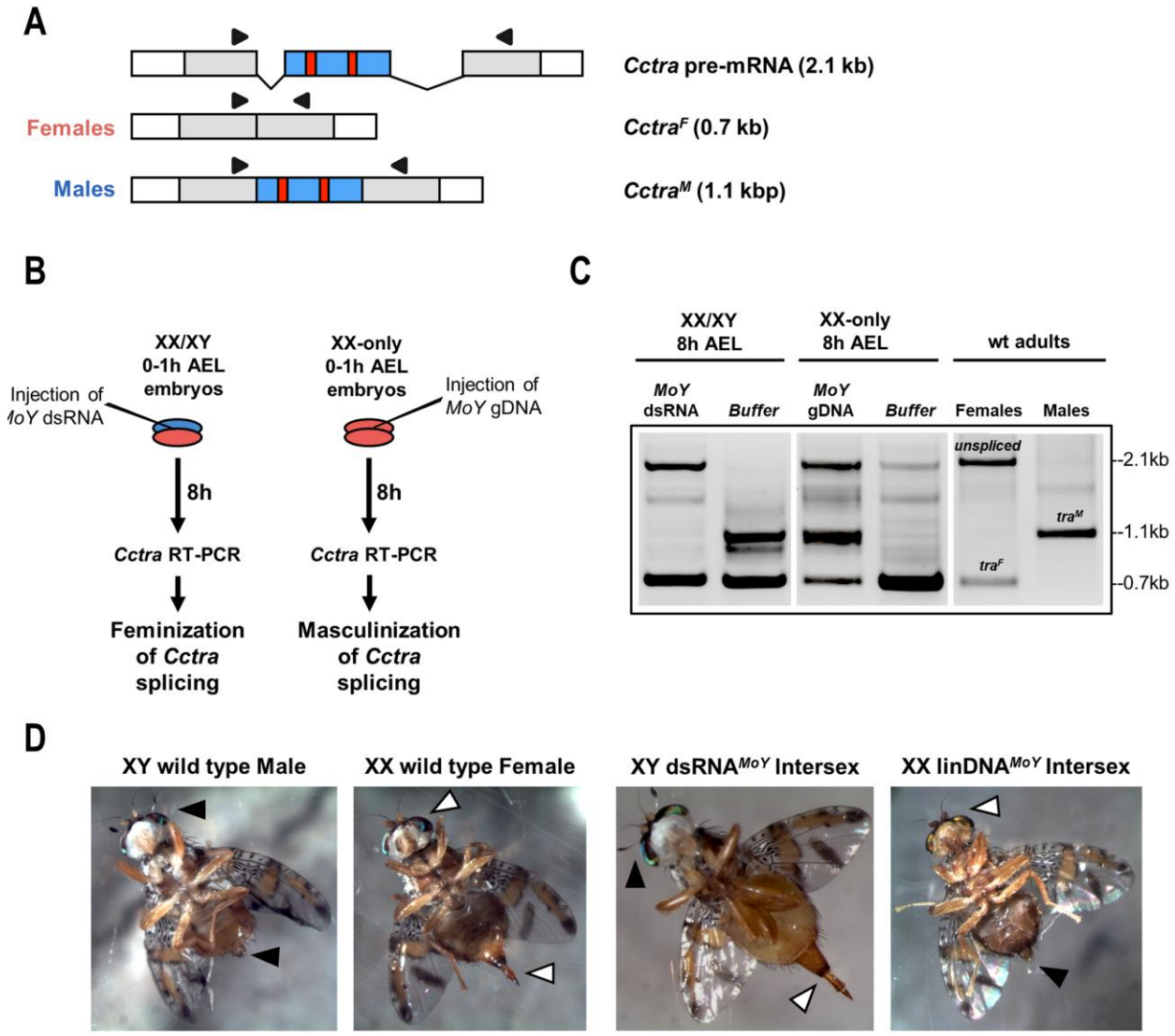

Fig. 2. *MoY* is necessary and sufficient for male sex determination. (**A**) Scheme of *Cctra* sex-specific transcripts. Coding regions, the male-specific exon and stop codons are shown in grey, in blue and red, respectively. Black arrows indicate primers for RT-PCR. (**B**) Schematic overview of embryonic *MoY* injection experiments and expected effects on *Cctra* splicing after 8h. (**C**) RT-PCR analysis showing splicing patterns of *Cctra* in 8h AEL embryos following injections. Left Panel: Transient *MoY*-eRNAi (at 0-1h AEL) depletes XX/XY embryos of the male-specific *Cctra* isoform at 8h AEL. Middle Panel: Injection of *MoY* genomic DNA (gDNA) is sufficient to instruct male-specific splicing of *Cctra* in XX-only embryos. The 2.1kb *Cctra* isoform is a female-specific unspliced transcript (*5*). Right Panel: wild type male and female flies for reference; black/white colours inverted in the gel photo for clarity. (**D**) Representative photos of Medfly wild type males,

females and intersexes. The XY intersex from *MoY*-eRNAi has a female ovipositor (white arrow) and a male-like head with orbital bristles (black arrow). The XX intersex from *MoY* linear genomic DNA (linDNA; PCR fragment) injections has male genitalia (black arrow) and a female-like head, without orbital bristles (white arrow).

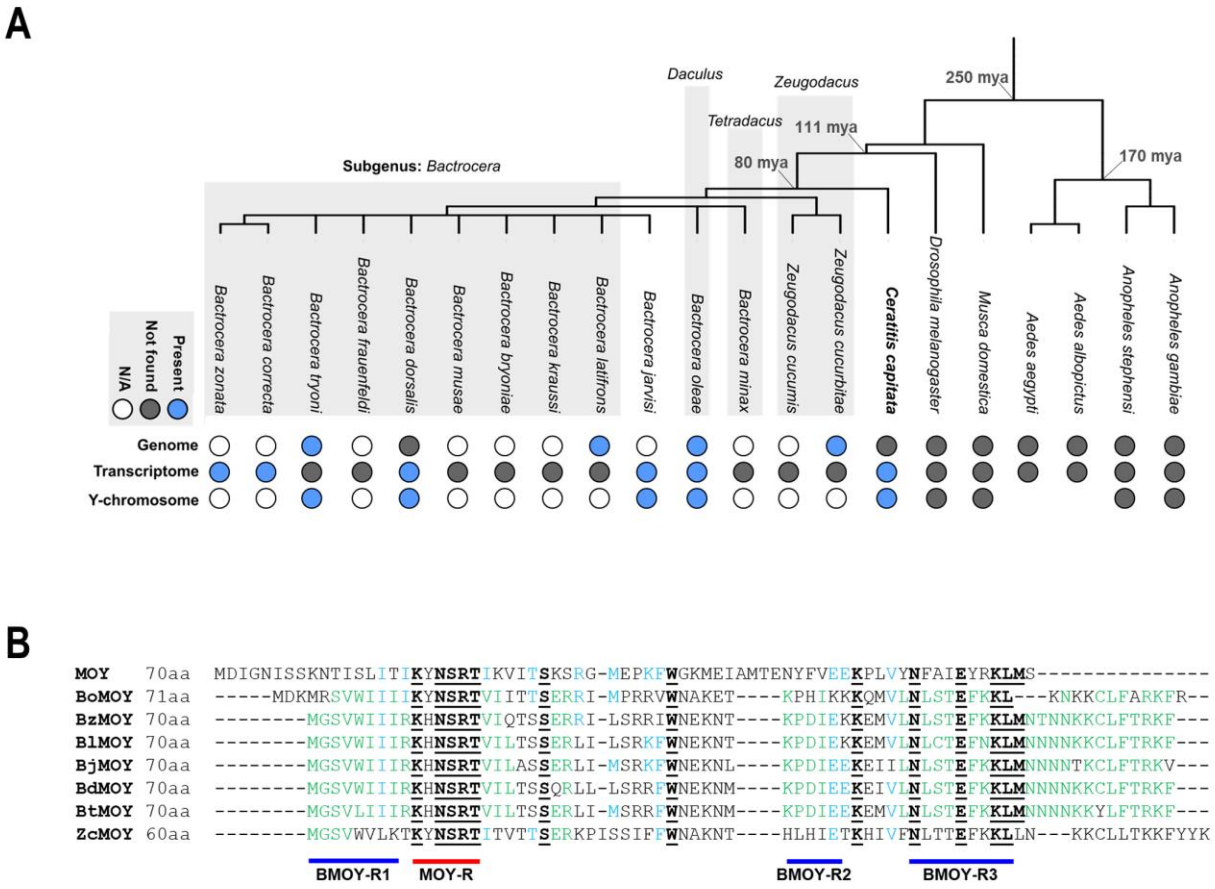

Fig. 3. *MoY* is conserved in Tephritidae species. (**A**) Phylogenetic tree of 14 Tephritidae species and 6 other dipteran species, in which primary sex determining signals (*M* factors in black and *Drosophila XSE* in red) have been molecularly characterized (*3*, *12–16*). Divergence times are indicated as million years ago (mya). *MoY* orthologues were discovered in 8 species, either in genome and/or transcriptome databases; their Y-linkage were assessed in the species where fly samples were available (see also Fig. 4A; supplementary text 5-12). N/A: not applicable. (**B**) Alignment of 8 MOY orthologous protein sequences including *C. capitata* (MOY), *B. oleae* (BoMOY), *B. zonata* (BzMOY), *B. latifrons* (BlMOY), *B. jarvisi* (BjMOY), *B. dorsalis* (BdMOY), *B. tryoni* (BtMOY), and *Zeugodacus cucurbitae* (ZcMOY). *B. correcta* (BcMOY) was not included because its sequence is not yet complete. Amino acids conserved in all 8 species are shown in black bold underlined; those conserved in *Ceratitis* and in at least one other species are shown in light blue and finally amino acids conserved in Tephritidae species other than *Ceratitis* are in green. A red line indicates a conserved short region, named as MOY-R (R= region), blue

lines indicate MOY conserved *Bactrocera* Regions, named as BMOY-R1, -R2 and -R3, respectively.

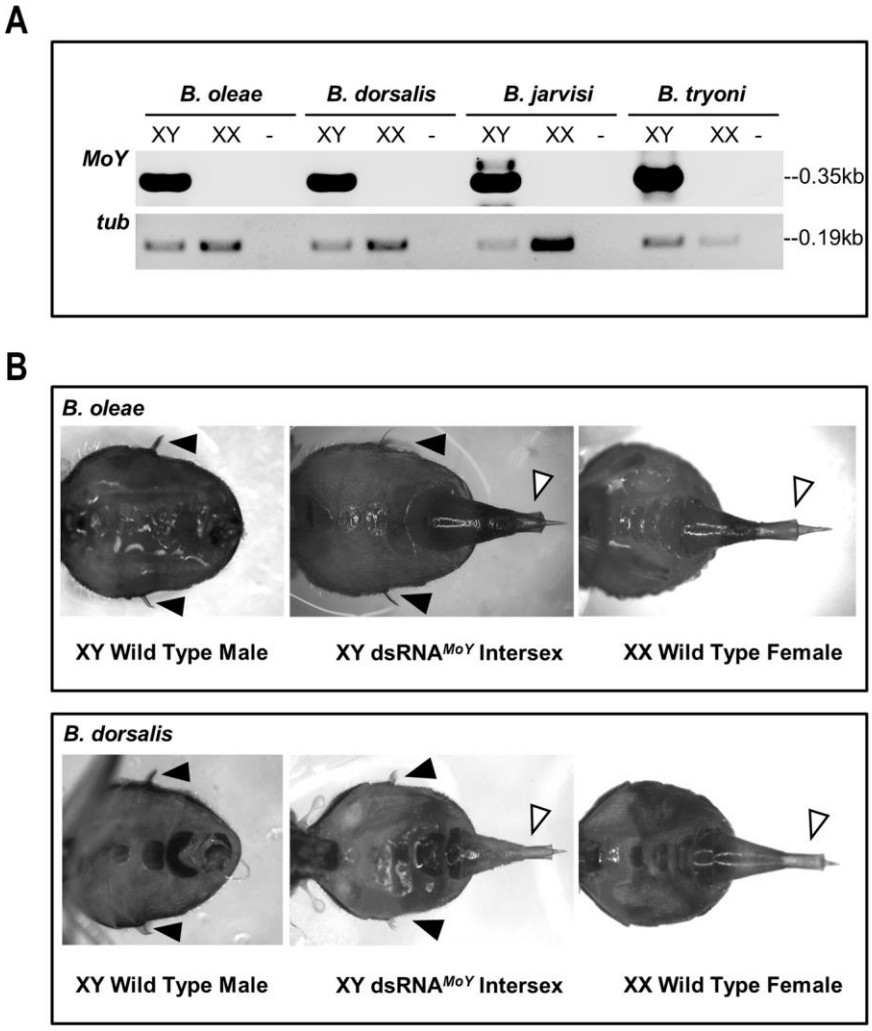

Fig. 4. *MoY* is Y-linked in *Bactrocera* species and functionally conserved. (**A**) PCR on male and female genomic DNA showing Y-linkage of *MoY* orthologues in *B. oleae*, *B. dorsalis*, *B. jarvisi*, and *B. tryoni*. Positive control: beta tubulin. Black/white colours are inverted for clarity. (**B**) *B. oleae* and *B. dorsalis* male and female wild type abdomens (left and right, respectively) and abdomens of intersex XY flies (middle), following *BoMoY*- and *BdMoY*-eRNAi (see also fig. S9). In both species, XY intersexes show the presence of female-specific (ovipositor; white arrow) together with male-specific characteristics (abdominal lateral bristles; black arrows).

Supplementary Materials for

Maleness-on-the-Y (*MoY*) orchestrates male sex determination in major agricultural fruit fly pests

Angela Meccariello*, Marco Salvemini*, Pasquale Primo, Brantley Hall, Panagiota Koskinioti, Martina Dalíková, Andrea Gravina, Michela Anna Gucciardino, Federica Forlenza, Maria-Eleni Gregoriou, Domenica Ippolito, Simona Maria Monti, Valeria Petrella, Maryanna Martina Perrotta, Stephan Schmeing, Alessia Ruggiero, Francesca Scolari, Ennio Giordano, Konstantina T. Tsoumani, František Marec, Nikolai Windbichler, Kallare P. Arunkumar, Kostas Bourtzis, Kostas D. Mathiopoulos, Jiannis Ragoussis, Luigi Vitagliano, Zhijian Tu, Philippos Aris Papathanos†, Mark D. Robinson† and Giuseppe Saccone†

*These authors contributed equally to this work

†Correspondence to: giuseppe.saccone@unina.it; mark.robinson@imls.uzh.ch; p.papathanos@mail.huji.ac.il

This PDF file includes:

Author contributions

GS conceived the production of XX-only embryos for mRNA-seq, rapid functional tests by RT-PCR (after injection of dsRNA or gDNA) in early embryos to identify the M factor, injections of recombinant MOY protein, the use of *Fam18* strain carrying a shorter Y, and identified *MoY* Tephritidae orthologues. GS, MDR, MS, ZT, BH, AK, JN, PAP, NW designed and performed bioinformatics analyses, with contributions from AM and PP. ZT and BH performed preliminary CQ analysis. MS, VP and AM prepared RNA and DNA for sequencing. MDR performed RNA sequencing, DE and CQ analyses, PacBio *Fam18* sequencing, and Canu assembly together with SS. MS performed transcriptome assemblies and developed a local web-tool with graphical interface for Tephritidae BLAST searches. AM, PP and GS selected *MoY* as a first *M* candidate from the list of putative male-specific transcripts by MDR and from the list of novel putative male-specific transcripts by ZT and BH (supplementary text S2). AM performed eRNAi, CRISPR/Cas9, and *MoY* DNA injections and RT-PCR analyses demonstrating *MoY* function. PP, AG, MAG, FF, DI and MMP maintained the strains, performed crosses and DNA/RNA molecular analyses. LV and AR purified recombinant MOY protein and performed structural/similarity analyses to protein databases. PP performed MOY protein embryos injections. FM and MD performed *in situ* hybridization of *MoY*. KDM, JR, KTT, MEG performed qRT-PCR analysis of *MoY* in *C. capitata* and *MoY* expression analysis in *B. oleae*. KM and JR provided *Bactrocera oleae* genome assembly data. AM, FS, PK and KB performed *MoY* RNAi on *MoY* orthologues of *B. oleae* and *B. dorsalis*. PK and KB performed molecular analysis on transformed XY females in these two species. GS, AM, MS and PAP prepared the figures. All authors discussed the data. MS, PAP, SMM, EG and LV provided essential reagents. GS wrote the manuscript with intellectual input from all authors, especially PAP, MDR, MS, and LV. AM and MS contributed equally to the work. GS initiated and supervised the project.

Data and materials availability

All data is available in the main text or the supplementary materials. The 12kb *MoY* sequence genomic region has been deposited in GenBank under accession number MK330842. Correspondence and requests for materials should be addressed to GS (giuseppe.saccone@unina.it, MDR, (mark.robinson@imls.uzh.ch), or PAP (p.papathanos@mail.huji.ac.il).

Materials and Methods

INSECT STRAINS

Insect rearing and strains

The following Medfly strains were used in this study: 1) *Benakeion* (developed by P. A. Mourikis, Benakeion Institute of Phytopathology, Athens, Greece); 2) *Fam18* (*8*) (see below); 3) *Vienna8* (*2*), which carries a marked Y chromosome harboring a translocated wild type allele of the white pupae (*wp*) mutation (*2*). All strains were reared under standard laboratory conditions at 25°C, 70% relative humidity (RH) and 12:12 h light–dark cycle. Adult flies were fed yeast/sucrose powder (1:2). Eggs were collected in trays filled with distilled water and transferred to larval food (Piccioni lab, Italy) after hatching. Pupae were collected and stored in dishes until eclosion.

The *B. dorsalis* Saraburi strain (Thailand) was maintained at the Insect Pest Control Laboratory of the Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, (IPCL, Seibersdorf, Austria) for 84 generations prior to use in these experiments. Flies were reared at 25 ± 1 °C, 60 ± 5% RH with a 12:12 h light–dark cycle and fed with a standard laboratory adult diet containing sugar and hydrolyzed yeast in a 3:1 ratio and water *ad libitum*. Eggs were collected and transferred to larval medium containing 28% wheat bran, 7% brewer's yeast, 13% sugar, 0.21% sodium benzoate, 0.21% nipagin, 1% HCl, 50% water.

The *B. oleae* Greece-Lab strain (Department of Biology, "Demokritos" Nuclear Research Centre, Athens, Greece) was maintained at the IPCL for 180 generations. Flies were reared at 25 ± 1 °C, 60 ± 5% RH with a 14:10 light-dark cycle and fed with a standard laboratory adult diet consisting of 75% sugar, 19% hydrolyzed yeast and 6% egg yolk powder. Eggs were collected and transferred to larval diet containing 550 mL water, 20 mL Extra Virgin Olive Oil, 7,5 mL Tween® 80 emulsifier, 0.5 g potassium sorbate, 2 g nipagin, 20 g sugar, 75 g brewer's yeast, 30 g soy hydrolysate, 4.5 mL HCl 36% and 275 g cellulose powder.

Description of the *Ceratitis capitata Fam18* strain

We used the *Fam18* strain (kindly provided by Dr. Gerald Franz, FAO/IAEA, Austria) for PacBio sequencing of male genomic DNA, due to the lower complexity of its shorter Y chromosome. This strain originated from a study by Willhoeft and Franz (*8*) who used cytogenetic approaches, including *in situ* hybridization with Y-specific repetitive probes (*21*) and Y-autosome translocations, to map the M-factor to the pericentromeric region of long arm of the *C. capitata* Y chromosome. This region represents approximately 15% of the entire Medfly Y chromosome. The *Fam18* Y-chromosome harbors an internal deletion of the long arm, but not a reciprocal Y-autosomal translocation (Franz, G., FAO-IAEA Pest Control Unit, person. comm. to GS, 2010). To verify at the genomic level that the *Fam18* Y chromosome is smaller compared to a wild type strain, we mapped using the Bowtie aligner (*22*) male and female Illumina reads from the *Fam18* strain and as a control the *ISPRA* wild type strain (male reads: SRR847687 and SRR847380; female reads: SRR847688 and SRR847689) against four previously described *Y*-specific genomic sequences (length 1-5kb) containing Y-specific repetitive elements (*8*, *23*). *Fam18* read counts are significantly lower for three of four Y-linked repetitive elements compared to read counts from the ISPRA strain (table S7). Conversely, the read counts on a single copy control gene (*Cctra-2;* (*24*) are comparable between the two strains.

SEQUENCING AND BIOINFORMATICS

Production of XX-only and XX/XY embryos for RNA sequencing

To generate the pool of XX-only embryos that were used as a reference as they lacked the Y-linked *M* factor, we microinjected wild type embryos with dsRNA targeting *Cctra*, as previously described (*5*). This produced also fertile sex-transformed XX males lacking a Y chromosome. These males were then individually crossed to 3 wild type virgin females each, to obtain embryos harboring no Y chromosome. XX males were selected from XY males by molecular karyotyping on each embryonic progeny. Embryos were collected 4-8h after egg laying (AEL) from each cross and total RNA was extracted for sequencing from 22 embryo clutches. We used RT-PCR on the Y-specific *Cclap* (*25*) to confirm that these embryos were composed entirely of XX-only individuals and RNA samples were pooled into 3 independent sets for sequencing. Total RNA was prepared also from mixed XX/XY embryos 4-8 h AEL from a cross of wild type *Benakeion* XX

and XY flies. We extracted RNA from three biological replicates using TRIzol reagent (Thermo Fisher Scientific), according to the manufacturer's protocol.

RNA sequencing, *de novo* transcriptome assembly and differential gene expression

The integrity and purity of extracted total RNA was evaluated using a NanoDrop 2000c (Thermo Fisher) and the Agilent 2100 Bioanalyzer with the RNA 6000 Nano kit (Agilent, Santa Clara, CA, US). All RNA samples had a A260/280 ratio higher than 2.1 and a RIN (RNA Integrity Number) value higher than 7.8. Total RNAs were depleted of ribosomal RNA with Ribo-Zero rRNA Removal kit (Epicentre). Libraries for sequencing, with an average insert size of 300bp, were generated with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). Illumina stranded and paired-end (PE) sequencing was performed at NXT-GNT (Belgium) yielding a total of 85,832,487 XX/XY and 94,085,461 XX-only PE 100bp-long reads (BioProject PRJNA434819 in NCBI SRA archive). A transcript catalogue was produced *de novo* using the combined XX-only and XX/XY embryos 4-8h Illumina read datasets, concatenated into two paired FASTQ files, and the Trinity assembler (*26*). The assembly was run with default parameters and **-**SS_lib_type RF -jaccard_clip -normalize_reads flags set, after adapter removal and quality trimming using Trimmomatic v.0.33(*27*) and depletion of mitochondrial and ribosomal reads, as described (*25*). Assembly statistics are presented in table S6. Transcript-level quantifications for each sample was done against this catalogue using Kallisto (*28*) (Data S2). Differential gene expression was performed using edgeR (*23*), with cut-off values of FDR (False Discovery Rate) < 0.05 and logFC (Fold Change) > 0 (Data S3). FDR = p-value adjusted for multiple testing with the Benjamini-Hochberg procedure.

Illumina short read WGS of *C. capitata Fam18* strain

Five adult males and five adult females from the *C. capitata Fam18* strain were collected and genomic DNA was extracted from the two sexed pools separately, as described below (*29*). Each pool was sequenced with 1 lane of Illumina HiSeq2500 resulting in 243,152,822 male and 248,886,022 female 2x126-bp PE reads, respectively (BioProject PRJNA435549 in the NCBI SRA archive).

Illumina short read WGS, trimming, error-correction and assembly

The Illumina raw short-read data from *Fam18* male and female samples had approximately 126X and 129X coverage, respectively (assuming a genome size of 485Mbp). Trimmomatic v0.35 (*27*) was used to remove Illumina adapters in paired-end mode, trimmed by quality values in maxInfo mode with target_length=0 and strictness=0.4 parameters, and selected for a remaining read length of at least 50 bases. The resulting trimmed reads were used for all purposes except for the error correction of the PacBio reads.

For the PacBio correction, we first corrected the trimmed short reads with BFC (*30*) (r181) using k-mer length=33 and genome size=485Mb as parameters. Meraculous (*31*) assembled those corrected short reads into unitigs with genome size=485Mb, num_prefix_blocks=8, and automatic k-mer length identification. To determine the library parameters for the corrected short reads, we mapped the reads to the Ccap01172013 genome from Baylor College7 using bowtie2 (*22*) v2.2.5 with a maximum insert length of 1000bp. This resulted in a mean insert size of 364bp with a Gaussian standard deviation of 79bp. The average read length of the corrected reads was 121bp. The corrected reads and the unitigs were then used to correct the PacBio reads.

PacBio WGS of *C. capitata Fam18* strain

Five adult males from the *C. capitata Fam18* strain were collected and genomic DNA was extracted using the Holmes-Bonner protocol (*29*). Extracted genomic DNA was resuspended in 10 mM Tris-HCl, pH 8.5, and its integrity and purity were assessed using the NanoDrop 2000c (Thermo Fisher Scientific) and 1% agarose gel electrophoresis, resulting in A260/280 ratio of 1.8. Genomic DNA was sheared by g-TUBE (Covaris) and a BluePippin instrument (Sage Science) was used to select fragments from 15-20kb in length. A large insert library was prepared using Pacific Biosciences recommended protocols and P6-C4 reagents. The resulting sample was sequenced at the Functional Genomics Centre Zurich (Switzerland) with 20 SMRT cells on a Pacific Biosciences RSII instrument. The PacBio SMRT cell reads have been submitted to the NCBI SRA archive with the accession number: BioProject PRJNA435534.

PacBio WGS error-correction and assembly

PacBio DNA sequencing using P6-C4 chemistry for 20 SMRT cells was conducted; extraction of sequences into FASTQ files was done by Dextractor (https://github.com/thegenemyers/DEXTRACTOR) with a minimum sequence quality parameter set to 400 produced ~32X genome coverage (assuming a genome size of 485Mbp) with an average read length of 5,526bp and a N50 length of 8,656bp, and a maximum length of 68,782bp. We combined all SMRT cells and then split the data into chunks of 400 Mb for cluster processing. The error correction was done by Proovread (*32*) using the corrected first reads from the PE short read data and the unitigs. We did not use the second reads as they were of lower quality and the additional coverage would have slowed down the error correction without improving the result. The only additional parameter set was coverage=62. This resulted in a total of ~19X corrected sequences with an average read length of 2,855bp, N50 length of 4,695bp, and maximum length of 33,053bp. This set of corrected reads was used to build a *de novo Fam18* genome assembly using Canu v1.0 with -pacbio-corrected set and genomeSize=0.45g. This resulted in 12,165 assembled contigs with a total length of 496Mb, an average read length of 40,769bp, N50 length of 62,964bp, and maximum length of 1.0Mbp.

Bactrocera spp. transcriptome assemblies and *MoY* orthologues search

Illumina RNA-seq data for 14 *Bactrocera* species were downloaded from NCBI SRA archive and *de novo* assemblies were produced using the Trinity assembler (*26*) with default parameters and -normalize_reads and -jaccard_clip flags set, as described (*33*). Assembly statistics and accession numbers of the utilized data set are shown in Data S4.

tBLASTn searches of the *de novo B. oleae* transcriptome, using MOY as protein query, identified a transcript encoding a putative protein with 36% identity over a 64 aminoacids (aa) long region (BoMOY, 70aa; MK165746) (supplementary text S5-S12). tBLASTn searches using BoMOY on *B. oleae* NCBI WGS database, failed to find the full-length corresponding *BoMoY* gene, but revealed the presence of 4 genomic sequences (Sequence IDs: JXPT01043932.1; LGAM01008500.1; LGAM01009849.1; LGAM01000393.1) encoding for truncated 15-41aa long *BoMoY* paralogous sequences and showing 37-100% aa identity to BoMOY protein. The identified *B. oleae MoY* orthologue was used as query to search by tBLASTn for MOY orthologues in other 13 assembled *Bactrocera* transcriptomes. *MoY* orthologues were found only in *B. jarvisi* (MK165748) and *B. zonata* (MK165752). We also searched for *MoY* orthologues by tBLASTn searches on the NCBI SRA and WGS databases. Sequences encoding putative full-length *MoY* orthologues were found in both *B. tryoni* (JHQJ01009763.1) and *B. latifrons*, showing 63% aa

identity with BoMOY (MIMC01001452.1). tBLASTn analyses also revealed the presence in *B. tryoni* of a shorter *MoY* paralogue (encoding a 50aa long BtMOY) in the same genomic contig, 4kb apart from the first *BtMoY* copy (encoding a 70aa long protein). The two *BtMoY* paralogues showed 90% protein identity over a 50aa long region. Similarly, in *B. latifrons* we found a truncated paralogue *BlMoY* gene in a different genomic contig (MIMC01001198.1), encoding for truncated 36aa long BlMOY sequence, with 78% identity to BlMOY and 63% to BoMOY. Finally, by tBLASTn searches we identified a SRA read from *B. dorsalis* transcriptome analysis data (SRR316210) encoding part of the putative BdMOY, corresponding to the N-terminus. We performed PCR on cDNA from 5-8h old embryonic RNA using primers *BdMoY+/BdMoY-* designed on the corresponding paired Illumina reads (table S8), leading to the isolation of the *BdMoY* full-length coding region.
We validated the *MoY* male-specific Y-linked genomic localization by amplification of genomic DNA of *B. oleae*, *B. dorsalis*, *B. tryoni* and *B. jarvisi* using the specific primers pair reported in table S8. The following list comprises the GenBank accession numbers of the 8 putative MOY orthologous protein sequences: BoMOY-A MK165746, BoMOY-B MK165747, BjMOY MK165748, BdMOY MK165749, BtMOY MK165750, BlMOY MK165751, BzMOY MK165752, RzMOY MK165753, ZcMOY MK165754.

CHROMOSOME *IN-SITU* HYBRIDIZATION

Probe preparation and labelling

A 1.5kb fragment of *MoY* was amplified from a plasmid by PCR with primers New_*MoY*_F1/New *MoY*_R (table S8). Probe labelling was performed using an improved nick translation procedure (*34*) with some modifications. The modified 40 µL reaction contained 500 ng of unlabeled DNA, 50 µM of each dATP, dCTP, and dGTP, 10 µM of dTTP, 20 µM digoxigenin-11-UTP (Roche Diagnostics, Mannheim, Germany), 1× nick translation buffer (50 mM Tris-HCl pH 7.5, 5 mM MgCl2, 0.005% BSA), 10 mM ß-mercaptoethanol, 1.25×10-4 U/µl DNase I and 1 U/µl DNA polymerase I (both ThermoFisher Scientific, Waltham, MA, USA). The reaction was incubated at 15 °C for 40 min and then purified using an illustra Sephadex G-50 column (GE Healthcare, Buckinghamshire, UK).

Chromosome preparation

Spread preparations of mitotic chromosomes were made from the brain (cerebral ganglia) of third instar larvae following a previously described method (*35*) with slight modifications. Briefly, the brain was dissected in a physiological solution, transferred into hypotonic solution (0.075 M KCl) for 10 min, and then fixed for 15 min in freshly prepared Carnoy's fixative (ethanol:chloroform:acetic acid, 6:3:1). The fixed tissue was spread in a drop of 60% acetic acid on the slide at 45 °C using a heating plate. Then the preparations were passed through a graded ethanol series (70%, 80%, and 100%, 30 s each), air dried, and stored at −20 °C. Before further use, the preparations were again dehydrated in the ethanol series, immediately after removal from the freezer.

Fluorescence *in situ* hybridization with tyramide signal amplification (TSA-FISH)

The TSA-FISH was performed according to published protocols (*35*, *36*) with some modifications. Slides were pre-treated with 200 µg/mL RNase A in 2× SSC at 37 °C for 1 h, with 0.01 M HCl at 37 °C for 20 min, 1% H_2O_2 in 2× SSC at room temperature (RT) for 30 min, and

5× Denhardt's solution (0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, and 0.1% bovine serum albumin in ultrapure water) at 37 °C for 30 min. The probe cocktail containing 50% deionized formamide, 10% dextran sulphate, and 0.5 ng/µL of the probe in 2× SSC was applied to chromosome preparation, covered with cover slip and denatured at 70 °C for 8 min. Hybridization was carried out at 37 °C overnight.

After hybridization, slides were washed three times in 50% formamide in 2× SSC at 46 °C for 5 min, three times in 2× SSC at 46 °C for 5 min, three times in 0.1× SSC at 62 °C for 5 min, and in TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween-20) at RT for 5 min. The slides were blocked in TNB buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5% Blocking Reagent; PerkinElmer, Waltham, MA, USA) at RT for 45 min. The probe was detected by anti-digoxigenin-POD (Roche Diagnostics) diluted 1:20 in TNB buffer. Tyramide amplification was carried out using the TSA™ Plus Fluorescein System (PerkinElmer) according to the manufacturer's instructions. The slides were incubated with tyramide working solution at RT for 20 min and then washed as previously described (*35*). Chromosomes were counterstained with 0.5 µg/mL DAPI (4′,6-diamidino-2-phenylindole; Sigma-Aldrich, St. Louis, MO, USA) in 1× PBS at RT for 15 min. Finally, the slides were mounted in antifade based on DABCO (1,4-diazabicyclo (2.2.2)-octane; Sigma-Aldrich).

Chromosome preparations were observed under a Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany). Black-and-white images were captured separately for each fluorescent dye with a monochrome CCD camera XM10 using cellSens Standard software version 1.9 (both Olympus Europa Holding, Hamburg, Germany). The images were pseudo-coloured and merged using Adobe Photoshop CS5 (Adobe Systems, San Jose, CA, USA).

FUNCTIONAL STUDIES

MoY gene structure and expression

BLASTn searches of the PacBio Canu assembly (*Fam18* strain, using *C. capitata MoY* 681nt long transcript sequence (including 5' and 3' UTRs; *Benakeion* strain) as DNA query (GenBank acc. num. MK165756), identified a highly similar corresponding genomic sequence (95% identity). The *MoY* putative coding region of the transcript is 99% identical in the PacBio genomic sequence, with only 2 SNPs, with the second inducing a conservative amino acid substitution at position 63 (I->M) (GenNank acc. num. MK165755).

To evaluate *MoY* expression, RNA was extracted from several developmental stages and tissues of Medfly. Per each embryonic sample, a pool of ~ 50 eggs were collected within 20 min AEL. In order to examine different time points throughout embryonic development, each pooled-sample was further incubated at rearing conditions until the desired developmental time was reached, and then transferred respectively in Extrazol (Nanogen Advanced Diagnostics, Turin, Italy). Each embryonic sample refers to the time after oviposition when embryos were collected in Extrazol (Nanogen Advanced Diagnostics, Turin, Italy). In total, 10 different embryonic time points (1h, 2h, 3h, 4h, 5h, 6h, 7h, 8h, 15h and 48h), 1st and 3rd instar larvae stages (10 larvae each sample), pupae (10 pupae per sample), adult flies and male dissected testes. All samples were collected in triplicates, except for males, testes and embryos at 15h and 48h, which were in duplicates. We found no detectable expression of *MoY* after 15h (Data S1).

Following RNA extraction, genomic DNA was removed using a DNase treatment with 1.0 unit of TURBO™ DNase (Invitrogen, CA) according to manufacturer's instructions. Total RNA was reverse transcribed using the PrimeScript RT reagent Kit (TaKaRa, Japan) according to

manufacturer's instructions. The resulting cDNA was used in the subsequent q RT-PCR reactions. Primers used are shown table S8. Expression levels were obtained relative to a housekeeping gene based on evaluating the reference genes list of each species (*37*). The *β-tub* gene was used to normalize expression in eggs, whereas *Rpl19* was used for larvae, pupae, male and female adults respectively (table S9).

Embryonic RNA interference

MoY, *BdMoY*, and *BoMoY* DNA fragments were PCR amplified using genomic DNA from a male adults of *C. capitata, B. dorsalis* and *B. oleae* (List of primers in table S8), introducing a T7 promoter sequence at each extremity. PCR was performed with Phusion® High-Fidelity DNA Polymerase (New England Biolabs) according to manufacturer instructions. The PCR product was purified with a phenol/chloroform purification and eluted in ddH_2O in a final concentration of 1 µg/µL. *In vitro* transcription of *MoY* dsRNA was performed using the Ambion MEGAscript® RNAi kit T7 RNA polymerase, following manufacturer instructions. The final product was checked on 1% agarose gel electrophoresis and quantified on NanoDrop 2000c. The dsRNA was eluted from the purification column using elution buffer, containing 5 mM KCl and 0.1 mM NaPO4, pre-heated at 98 °C, to get a final dsRNA yield of 1.5 µg/µL.

Embryos were collected 45 min AEL, placed on double-stick tape, hand-dechorionated, dehydrated in a Petri dish containing calcium chloride and covered with Halocarbon oil 700 (Sigma, H8898). A solution of 1 µg/µL dsRNA was microinjected for all species.

CRISPR and RNP complex assembly in *C. capitata*

sgRNA was designed using CHOPCHOP (*38*). Production of the sgRNA template was performed, with minor modifications, as described (*11*), with CRISPR-*MoY*-F and the invariant reverse primer (PAGE-purified, Invitrogen) (Table S8). sgRNA was synthesized according to instructions of the Megascript® T7 kit (Ambion) with 1 µg of template and a 5' flanking T7 promoter as starting material. After RNA synthesis, template was removed by incubating with TurboDNase® (Ambion) for 15 min at 37 °C. Cas9 was expressed as his-tagged protein and purified from bacteria as described (*11*). Prior to injection, the RNP complex was prepared by mixing 1.8 µg of purified Cas9 protein with approximately 200 ng of sgRNA in a 5 µL volume containing 300 mM KCl (*11*). The mix was incubated for 10 min at 37 °C. A glass needle was filled with the pre-loaded sgRNA-Cas9 mix and injection was performed into the posterior end of embryos collected 45 min AEL, as described for the RNA interference in *C. capitata*. 48h after the injections, the slides were inspected for hatched larvae under a stereomicroscope. Hatched larvae were transferred to Petri dishes containing larval food.

RNA extractions and RT-PCR

Total RNA was extracted from pools of embryos, larvae or adults using TRIzol Reagent (Thermo Fisher Scientific) following the manufacturer's instructions. Oligo-dT-primed cDNA was prepared from DNase I-treated total RNA using EuroScript® m-MLV reverse transcriptase (Euroclone®). RT-PCR expression analysis was performed using the primers listed in table S8.

Molecular karyotyping of *Bactrocera* adults

DNA was extracted using EXTRACTME DNA TISSUE kit (EM03) following manufacturer's instructions. PCR was performed using Qiagen Taq PCR Master Mix Kit (201443) according to manufacturer instructions. The PCR amplifications of Y-specific fragments were

performed with primers *BdMoY*-F/*BdMoY*-R for *B. dorsalis* and *BoMoY*-F/ *BoMoY*-R for *B. oleae*. PCR analysis was performed using primers described in table S8.

Amplification and cloning of the 5kb fragment of *MoY*

PCR amplification of a 5kb long *MoY* genomic fragment (Fig. 1C) of was done with the LongAmp® Taq DNA Polymerase following manufacturer's instructions in a final volume of 100 µL using the primers Gen_CcMoY_F and Gen_CcMoY_R (table S8). The 5kb *MoY* fragment was then cloned in the pGEM®-T Easy vector (Promega) following manufacturer's instructions. The PCR product and the plasmid were purified with phenol:chloroform (1:1), precipitated in ethanol and resuspended in injection buffer (containing 5 mM KCl and 0.1 mM $NaPO_4$) for a final concentration of 1 µg/µL and injected in *C. capitata* embryos 0-1h AEL.

MOY protein expression, purification and characterization

The *MoY* gene was amplified from male *C. capitata* genomic DNA with two primers carrying the *Nco*I and *Xho*I restriction sites: CcOrf2fw (*Nco*I)/ CcOrf2rv (*Xho*I) (Table S8). The fragment was then cloned into the pETM-13 expression vector. The resulting gene contained an extra C-terminal tag of six histidine residues. The MOY protein was then heterologously expressed in *E. coli* BL21 (DE3) cells. MOY was purified from inclusion bodies after removing the soluble proteins by sonicating the bacterial resuspension in a native buffer. The pellet was then resuspended in a denaturing buffer containing 8 M Urea, 50 mM Tris-HCl and 200 mM NaCl (pH 8.0) and left at room temperature for 2 hours to facilitate its solubilization. After the centrifugation, the supernatant was loaded on a Ni-NTA resin (Qiagen) equilibrated with the denaturing buffer. The protein was washed with ten volumes of the denaturing buffer and then eluted with a solution containing a high concentration of imidazole (150-300 mM pH 8.0). The homogeneity of the eluted protein was evaluated by SDS–PAGE analysis. MOY protein was refolded in solution containing 20 mM ethanolamine and 0.5 M L-arginine (pH 9.0-10.0) using rapid dilution (1:50 v/v) with three additions of the same volume over a 24h period. The protein was concentrated using Amicon centrifugal devices (Millipore) up to a concentration of 70 µg/mL. The molar mass of the refolded MOY was determined by mass spectrometer (ESI-TOF) (9250.28 Da) in agreement with that computed from the tagged sequence (9251.7 Da).

Supplementary Text

Supplementary Text S1

Identification of *MoY*, a single copy intronless gene

By filtering for transcripts predicted to be on the Y chromosome based on a CQ analysis and expressed specifically in the mixed XX/XY embryos but absent in the XX-only embryos, we selected 19 candidate transcripts, corresponding to 10 distinct transcriptional units (Fig. 1A; table S2). Of these 19 transcripts, 7 were excluded as *M*-factor candidates, because they did not map to the *Fam18* PacBio male genome assembly (table S1). 11 out of the remaining 12 transcripts were considered of minor interest because they showed similarity to multiple paralogous sequences in the *Fam18* genome. We also investigated if any of these 12 transcribed sequences showed similarity by BLASTn to XY but not XX embryonic transcripts in *B. oleae* and selected 3 transcripts as being specific to XY samples (table S1). Only 1 of these 3 transcripts (DN40292_c0_g3_i1) mapped to a single location by BLASTn to the *Fam18* Canu assembly (contig00013010, length=12420bp) indicating that it may be single copy (cutoff e-value=0).BLASTn analysis also identified 4 shorter *MoY* related sequences within the 12kb long contig 00013010, showing 70% identity over 200-500bp (see the 4 dashed red lines in Fig. 1C) and 2 shorter antisense RNAs (see 2 red inverted arrows in Fig. 1C) overlapping the *MoY* open reading frame. When BLAST hits cutoffs were relaxed, 18 additional contigs of the Canu assembly contained MOY-related hits. However, all hits correspond only to truncated versions (20-40aa long, showing 50-70% aa identity). tBLASTn searches using the MOY amino acid sequence against the NCBI *Ceratitis* reference genome or transcriptome (refseq_RNA) did not result in significant hits, except an unplaced genomic scaffold (NW_019377179.1) harboring a short, truncated *MoY* sequence (79% aa identity over a 19aa long region). BLASTn, tBLASTn and BLASTp searches at NCBI (either all species, Insecta or Diptera) using the *MoY* DNA or MOY amino acid sequence did not result in significant hits.

Supplementary Text S2

A preliminary screen for *M*-factor candidates

In previous attempts we tried to identify the Medfly *M*-factor using the CQ approach and publicly available data (RNA-seq: SRX272876 and SRX272878; male WGS: SRX276046, SRX275788 and SRX272878; female WGS: SRX275787, SRX276048 and SRX276047). CQs were calculated for *de novo* assembled transcripts from a combined embryonic (0-48 h old) and adult male RNA-seq datasets (SRX272876 and SRX272878). This initial attempt, which relied on limited RNA-seq data lacking biological replicates and suboptimal developmental staging of 0-48 h old embryos, led to the identification of 7 Y-linked male-specific transcriptional units (see Table below). 4 of these 7 were confirmed to be Y-linked by PCR on gDNA. 5 of the 7 had no hits in the NCBI Baylor assembled genome, as expected. BLASTx analyses of 5 genes on *C. capitata* and *D. melanogaster* protein databases showed that they likely correspond to transcribed genes encoding proteins with low similarity to sequences at NCBI. The remaining 2, *corvus* and *dorado* (Table below), showed neither protein similarity to sequences in these protein databases nor did they appear to have paralogous transcripts in the *de novo* transcript assembly. *MoY* was subsequently re-discovered as *corvus*.

Name	Hits on Medfly Baylor Genome	Hits to transcriptome assembly	BLASTx Ceratitis	BLASTx Drosophila
Orion	No	18 contigs (at least 7e-59)	Cytosol aminopeptidase-like, myb-like protein 1	Sperm-leucylaminopeptidase 3, isoform C
Lyra	Yes	6 contigs (at least 2e-44)	Putatitve gustatory receptor 59f	None
Aries	No	4 contigs (at least 4e-63)	Cytosol aminopeptidase-like	Sperm-leucylaminopeptidase 3, isoform C
Dorado	Yes	None	None	None
Pavo	No	4 contigs (at least 2e-49)	Cytosol aminopeptidase-like	Sperm-leucylaminopeptidase 3, isoform C
Norma	No	1 contigs (at least 1e-82)	NADH dehydrogenase (ubiquinone) chain I (mitochondrion)	NADH dehydrogenase subunit I (mitochondrion)
Corvus	No	None	None	None

Supplementary Text S3

Other novel transcriptional units in the *MoY* genomic flanking regions

BLASTn searches of the *de novo* embryonic transcriptome using contig00013010 (GenBank: MK330842) as probe were performed (with a 50% identity cut-off). 20 predicted transcripts, corresponding to 10 different genes, were mapped along contig (see Table below for details and Fig.1C: *MoY* is indicated by a red arrow (#8), 2 *MoY* antisense RNA (red arrow #9, #10), and weakly related *MoY* sequences (red dashed lines). Violet arrows (#4) and green arrows (#5) represent 5 copies of varying length of a sequence transcribed from both strands (DN40292 and DN40516); in light brown 3 (#1-3), brown (#6), light violet (#7) other unrelated transcripts. The color code and number in the table below match those of Fig. 1C.

N°	Name	Lenght	Position in PacBio contig00013010	BLASTn XX embryos	BLASTn NCBI hits	BLASTx NCBI hits	BLASTn Bo XY embryos	BLASTn Bo XX embyos	GenBank Acc. Num.
1	TRINITY_DN26767_c0_g1_i1	381 bp	552-258	85% in 212 bp	Yes	Yes (homeobox)	Yes	Yes	MK330832
2	TRINITY_DN45758_c0_g1_i1	212 bp	1330-1119	77% in 70 bp	None	Yes (phospholipase)	None	None	MK330834
3	TRINITY_DN6507_c0_g1_i1	244 bp	1687-1444	93% in 91 bp	None	None	None	None	MK330833
4	TRINITY_DN40292_c0_g1_i9	2865 bp	4837-6083 (88%) 6098-7025 (98%) 8721-9610 (93%) 9609-11390 (97%) 12145-12420 (96%)	83% in 169 bp	Yes	None	None	None	MK330838
5	TRINITY_DN40516_c0_g1_i5	1990 bp	6083-4443 (88%) 12420-11791 (92%) 9610-8323 (95%) 10244-9609 (96%)	None	None	None	None	None	MK330839
6	TRINITY_DN32944_c0_g1_i1	467 bp	8291-8747 (96%) 4443-4861 (96%)	none	None	None	None	None	MK330840
7	TRINITY_DN38978_c0_g1_i1	535 bp	6854-7338	74% in 350 bp	Yes	None	None	None	MK330841
8 - *MoY*	TRINITY_DN40292_c0_g3_1	681 bp	3620-4309	None	None	None	Yes	None	MK330835
9	TRINITY_DN77369_c0_g1_i1	243 bp	3760-3518	None	None	None	None	None	MK330837
10	TRINITY_DN104942_c0_g1_i1	285 bp	4263-3970	None	None	None	None	None	MK330836

Supplementary Text S4

Discovery of the *MoY* orthologue in the olive fly *Bactrocera oleae*

BLASTn and tBLASTn analyses with *MoY* in the XY-only and XX-only embryonic transcriptomes of *Bactrocera oleae* (assembled respectively from SRA SRX265053 and from SRX265052) identified a *MoY* orthologue only in XY samples, showing 77% nucleotide identity over a 57bp long region and 57% protein similarity over 67aa long MOY region. On the contrary, the sequences of 2 antisense *MoY* RNAs (corresponding to the *MoY* 5' and 3' UTRs) and all other transcripts present in the Canu contig containing *MoY* region were not found in these searches. BLASTp analysis showed that MOY and BoMOY share 63% aa similarity over a 58aa long region. We found *BoMoY* gene within a genomic scaffold assembled from WGS data (Sequence ID: JXPT01043932.1). Unlike *MoY*, *BoMoY* contains 2 introns, the first being 186bp long in the 5' UTR and a second 941bp long within the ORF region. *BoMoY*-specific PCR on sexed *B. oleae* genomic DNA confirmed that *BoMoY* is Y-chromosome linked (Fig. 4A).

Supplementary Text S5

Discovery of the *MoY* orthologue in the Australian pest *Bactrocera jarvisi*

BLASTn analysis with MOY and BoMOY protein sequence in XY-specific and XX-specific embryonic transcriptomes (embryos 3-5h AEL) of *Bactrocera jarvisi* (assembled from SRX697428, SRX697431, SRX697434, SRX697435) revealed the presence of a BjMOY orthologous sequence only in XY embryonic transcriptome. The predicted BjMOY protein is 70aa long and shows by BLASTp 76% aa similarity to BoMOY and 60% to MOY (Fig. 3B) (see below). We initially noticed that in all three isoforms a stop codon is wrongly predicted downstream of the start codon. Manual inspection of the SRA files allowed us to manually curate the sequence and replace the stop codon with one encoding serine, and we considered this as the functional and corrected sequence, being conserved with respect of BoMOY orthologue. The current lack of other *B. jarvisi* genomic data, particularly WGS, hampers our ability to confirm this.

Supplementary Text S6

Discovery of the *MoY* orthologue in the Oriental fruit fly *Bactrocera dorsalis*

tBLASTn searches of the oriental fly *Bactrocera dorsalis* SRA NCBI databases, using *BjMOY* amino acid sequence as a query, led to the discovery of SRA sequence, showing very high aminoacid sequence identity. PCR on male and female genomic DNA of *B. dorsalis*, confirmed that this sequence is indeed Y-specific (Fig. 4A). A cDNA fragment encoding a full length BdMOY ORF (70aa) was subsequently amplified from embryonic *B. dorsalis* total RNA, using 2 primers designed on the forward and reverse SRA sequence (SRA: SRR316210.7953824.1 and SRA: SRR316210.7953824.2; Table S8). Based on the full sequence, BLASTp analyses revealed that BdMOY is similar to BjMOY (70aa) (97% aa overall similarity), to BoMOY (80% aa overall similarity) and to MOY (60% aa similarity over a 55aa long region).

Supplementary Text S7

Discovery of the *MoY* orthologue in the Queensland fruit fly *Bactrocera tryoni*

tBLASTn using BdMOY on available WGS data of the *Bactrocera tryoni* (Qfly, Queensland fly) led to the discovery of a 7kb long scaffold (GenBank: JHQJ01009763.1) containing an ORF with 97% protein similarity to BdMOY. A second high related *BtMoY* paralog (94% nt identity; *BtMoY-2*) is also present 3.5kb downstream of the first *BtMoY* but appears to be truncated (55aa long). *BtMoY* shares 85% nt sequence identity over a 700nt long region with *BjMoY* of Australian species *Bactrocera. jarvisi*. The 2 MOY orthologous proteins share an overall 94% aa sequence similarity.

Supplementary Text S8

Discovery of the *MoY* orthologue in the melon fruit fly *Zeugodacus cucurbitae* (previously known as *Bactrocera cucurbitae*)

tBLASTn searches with BdMOY on available WGS database of *Zeugodacus cucurbitae* led to the discovery of a 1.3kb scaffold (GenBank: JRNW01040954.1) containing an ORF with 66% aa similarity over 24aa long N-terminal region and 84% over a 19aa long central region. A frameshift was observed within the *ZcMoY* putative genomic region, possibly due to various reasons: 1) sequencing error, 2) presence of a small intron, or 3) existence of multiple *ZcMoY* copies with some inactivated by mutations. The two *ZcMoY* putative coding regions encoded 30aa long and 43aa long ORFs, respectively. A BLASTp alignment of the 2 ZcMOY coding sequences with BdMOY allowed us to manually join them into a 54aa long putative ZcMOY. No corresponding *ZcMoY* transcript sequences were found in the available *Z. cucurbitae* SRA databases. BLASTp analysis showed that ZcMOY (54aa) is similar to BdMOY (70% aa overall similarity) and to MOY (60% aa similarity over a 55aa long region).

Supplementary Text S9

Discovery of the *MoY* orthologue in *Bactrocera latifrons*

tBLASTn searches with BdMOY aa sequence on available WGS database of *Bactrocera latifrons* led to the discovery of a 30kb long scaffold (Sequence ID: MIMC01001452.1) containing a putative ORF with an overall 90% protein similarity over a 70aa long region (BlMoY). A second scaffold (Sequence ID: MIMC01001198.1) was also identified containing a shorter *BlMoY* ORF (*BlMoY*-2), showing 91% protein similarity over a 36aa long region. tBLASTn searches of 5 available SRA databases from *B. latifrons* (adult males: SRX1007577; adult females: SRX1007576; embryos: SRX1007578; larvae: SRX1007579; pupae: SRX1007580) failed to find reads from *BlMoY* transcripts.

Supplementary Text S10

Discovery of the *MoY* orthologue in the peach fruit fly *Bactrocera zonata*

tBLASTn searches with BdMOY aa sequence on available SRA databases of *Bactrocera zonata*, led to the discovery of a number of reads in adult males and pupae (SRX2016848, SRX2016847), but not female or embryo samples (SRX2016849, SRX2016846). tBLASTn searches using BdMOY protein against a *de novo B. zonata* Trinity transcriptome assembly resulted in the identification of a 0.9kb long predicted transcript encoding for a 70aa long protein (BzMOY) and

showing 93% protein sequence similarity. BLASTp analyses showed that BzMOY is similar to BdMOY (92% aa overall similarity) and to MOY (56% aa similarity over a 60aa long region) (see below).

Supplementary Text S11

***MoY* orthologue in *Bactrocera correcta*, distributed in Southeast Asia**

A tBLASTn search of *Bactrocera correcta* pupal and adult male RNA sequence databases (SRA SRX2013590 and SRX2013591) led to the discovery of 2 partially overlapping reads, which encode a 35aa sequence, bearing 94% similarity to the C-terminus of BtMOY.

Supplementary Text S12

Analysis of the biophysical/structural properties of *Ceratitis capitata* MOY protein and its orthologues

Protein	Residues	Mw	Theoretical Isoelectric point
MoY	70	8186	9.5
BoMoY	71	8708	11.4
BzMoY	70	8467	10.8
BlMoY	67	7978	9.9
BjMoY	70	8426	10.2
BdMoY	70	8468	10.5
BtMoY	70	8492	10.3
ZcMoY	60	7360	10.1

Biochemical properties of MOY proteins. Due to the high pI value, all these proteins are positively charged at neutral pH. Therefore, they are potentially able to interact with negatively charged nucleic acids.

```
MoY         MDIGNISSKNTISLITIKYNSRTIKVITSKSRGMEPKFWGKMEIAMTENYFVEEKPLVYN      60
ZcMoY       --------MGSVWVLKTKYNSRTI----------TYFFWNAKNTH----LHIETKHIVFN      38
BoMoY       -----MDKMRSVWIIIIKYNSRTVIITTSERRIMPRRVWNAKETK----PHIKKKQMVLN      51
BjMoY       --------MGSVWIIIRKHNSRTVILTSSERLIMSRRFWNEKNMK----PDIEEKEIILN      48
BtMoy       --------MGSVLIIIRKHNSRTVILTSSERLIMSRRFWNEKNMK----PDIEEKEMVLN      48
BdMoY       --------MGSVWIIIRKHNSRTVILTSSQRLLLSRRFWNEKNMK----PDIEEKEIVLN      48
BlMoY       --------MGSVWIIIRKHNSRTVILTSSERLILSRKFWNEKNTK----PDIEKKEMVLN      48
BzMoY       --------MGSVWIIIRKHNSRTVIQTSSERRILSRRIWNEKNTK----PDIEKKEMVLN      48
                      :: ::  *:****:             .*.  :        :: * :: *
Prediction            eeeeeeee   eeeeee                    eee   eeeee


MoY           FAIEYRKLMS---------------70
ZcMoY         LTTEFKKLL---NKKCLLTKKFYYK60
BoMoY         LSTEFKKLK---NKKCLFARKFR--71
BjMoY         LSTEFKKLMNNNNTKCLFTRKV---70
BtMoy         LSTEFKKLMNNNNKKYLFTRKF---70
BdMoY         LSTEFKKLMNNNNKKCLFTRKF---70
BlMoY         LCTEFNKLMNNNNKKCLFTRKF---70
BzMoY         LSTEFKKLMNTNNKKCLFTRKF---70
              :  *:.**
Prediction    ehhhhhhhh       eee
```

Secondary structure prediction was performed using Prediction protocol as implements in PROMALS3D, which performs the prediction using multiple sequences. Helices and β-structured regions are denoted with e and h, respectively. As shown, the analysis of the CLUSTAL multiple alignments indicates that there are 11 conserved residues in all sequences. The most conserved region is located in the N-terminal portion of the proteins. Of relevance is the hexapeptide KXNSRT that contains two strictly conserved positively charged residues. The lack of sequence similarities with any protein with a known three-dimensional structure makes the determination of *MoY* putative structural properties difficult. Nevertheless, the reliability of *ab-initio* secondary structure predictions methods does provide some structural information. PROMALS3D predicts a significant level of secondary structure that for all proteins. Indeed, approximately 60% of the residues of these proteins are embodied in secondary structure elements. This value is in line with that observed for globular compact proteins. The inspection of the table reporting the pairwise alignments indicates that the most distant sequence of this ensemble is *MoY*. BjMOY, BdMOY, and BtMOY are very similar (97% similarity).

Pairwise sequence identities (%). The numbers in parenthesis represent the similarity (%)

	MOY	ZcMOY	BoMOY	BzMOY	BlMOY	BjMOY	BdMOY	BtMOY
MOY	-	28 (57)	22 (41)	21 (41)	23 (41)	22 (44)	22 (44)	23 (44)
ZcMOY	-	-	51 (70)	50 (70)	50 (69)	49 (68)	51 (69)	47 (65)
BoMOY	-	-	-	58 (64)	49 (57)	51 (60)	51 (62)	53 (60)
BzMOY	-	-	-	-	84 (86)	84 (90)	87 (93)	87 (90)
BlMOY	-	-	-	-	-	83 (89)	84 (91)	83 (87)
BjMOY	-	-	-	-	-	-	91 (97)	91 (94)
BdMOY	-	-	-	-	-	-	-	91 (97)
BtMOY	-	-	-	-	-	-	-	-

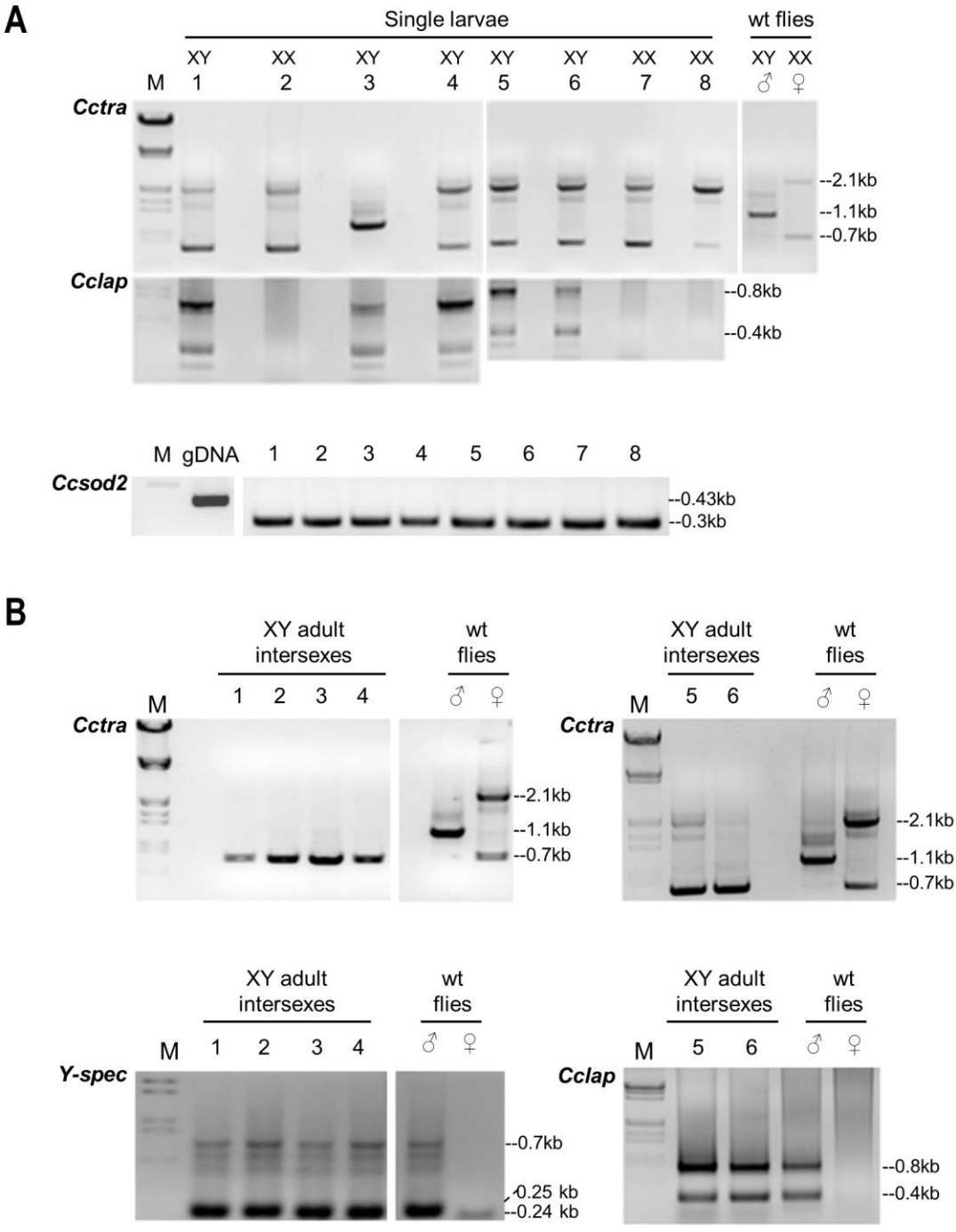

Fig. S1. Transient embryonic silencing of *MoY* leads to *Cctra* female-specific splicing in XY larvae and XY adults. (**A**) RT-PCR analyses of *Cctra* in single larvae that hatched from *MoY*-eRNAi embryos: 4 out of 5 XY individuals displayed female-specific splicing of *Cctra* (lanes 1,

4, 5 and 6). No effect was observed on female-specific *Cctra* splicing pattern in the 3 XX larvae (lanes 2, 7 and 8) and on male-specific splicing pattern in one XY larva (lane 3; eRNAi escaper). As a reference, *Cctra* cDNA fragments were amplified from wild type adult flies. Larvae were molecularly karyotyped for sex, using a Y-chromosome derived *Cclap* transcript (*25*); *Ccsod2* (*39*, *40*) (*C. capitata superoxide dismutase2*) was used as a cDNA positive control and as negative control for genomic DNA contamination (presence of genomic DNA would result in 0.43kb band, *Ccsod2* cDNA amplification would result in 0.3kb band: see gDNA lane). (**B**) RT-PCR analyses of *Cctra* splicing on 5 XY intersex flies revealed predominantly female-specific transcripts, indicating partial feminization of XY individuals. Molecular karyotyping was performed by RT-PCR using the Y-chromosome specific *CcYF/CcYR* primers (*9*) (lanes 1-4) or *Cclap* (lanes 5-6).

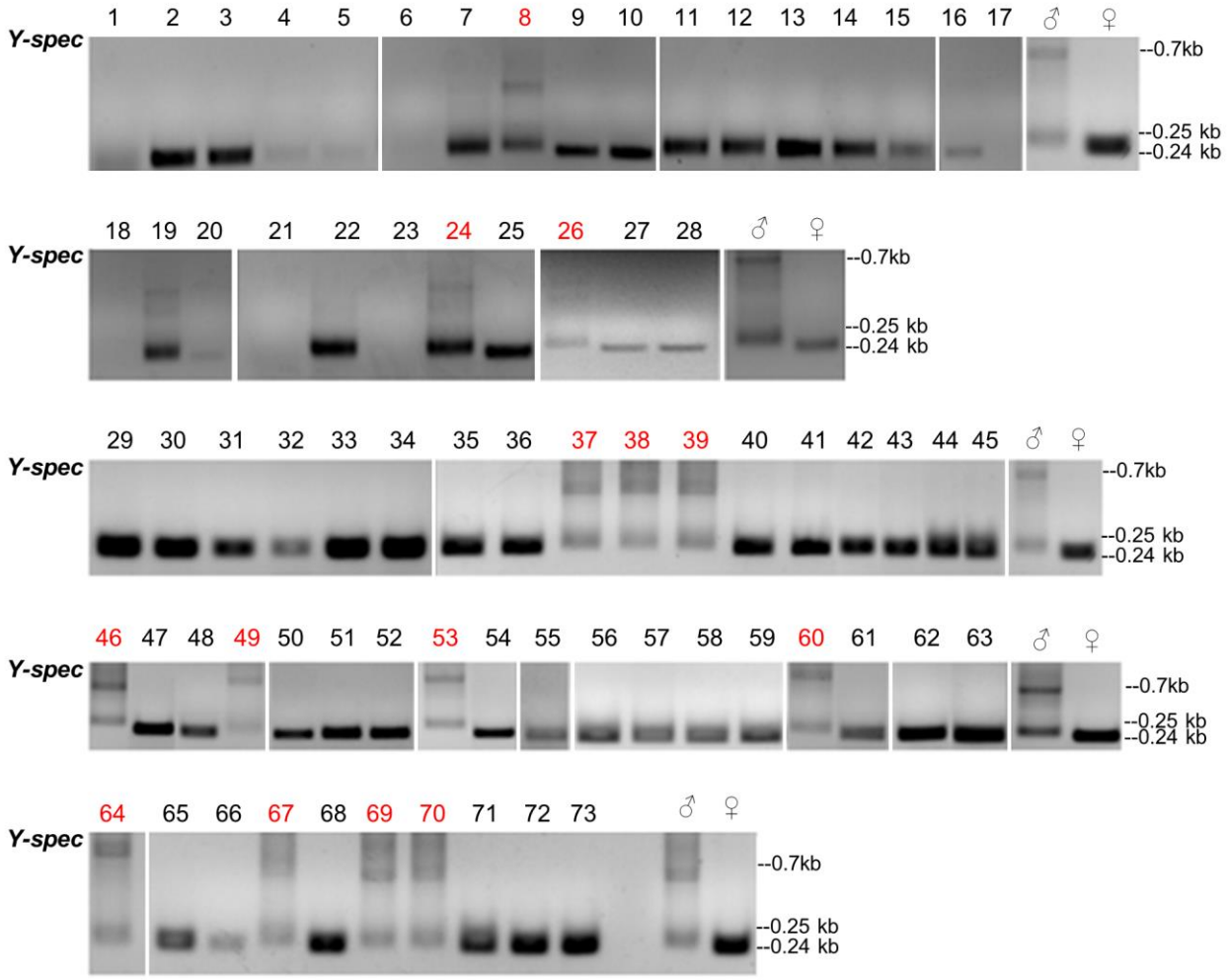

Fig. S2. Molecular karyotyping of G_0 adult females obtained from embryonic *MoY* RNAi: 73 adult females were molecularly karyotyped by PCR amplification from genomic DNA extracted from a small wing fragment (females 1-28) or from entire flies after crossing (females 29-46; see table S4), using the Y-chromosome specific *CcYF/CcYR* primers (*9*) (tables S2 and S4). Presence of the Y chromosome is indicated by a 700bp and 250bp amplicon (Y-spec) (*9*) which was detected in 14 out of 73 adult females (shown in red). In the remaining 59 adult females, only a smaller band was detected indicating the absence of the Y chromosome.

Fig. S3. Experimental diagram of the crossing strategy between transformed XY and transformed XX flies: (**A**) XY females obtained by transient *MoY* repression by RNAi transmitted the Y chromosome to the G_1 progeny determining male sex; (**B**) XY females obtained by *MoY* loss of function mutations by CRISPR/Cas9 transmitted the non-functional Y-linked *MoY* alleles to XY G_1 females.

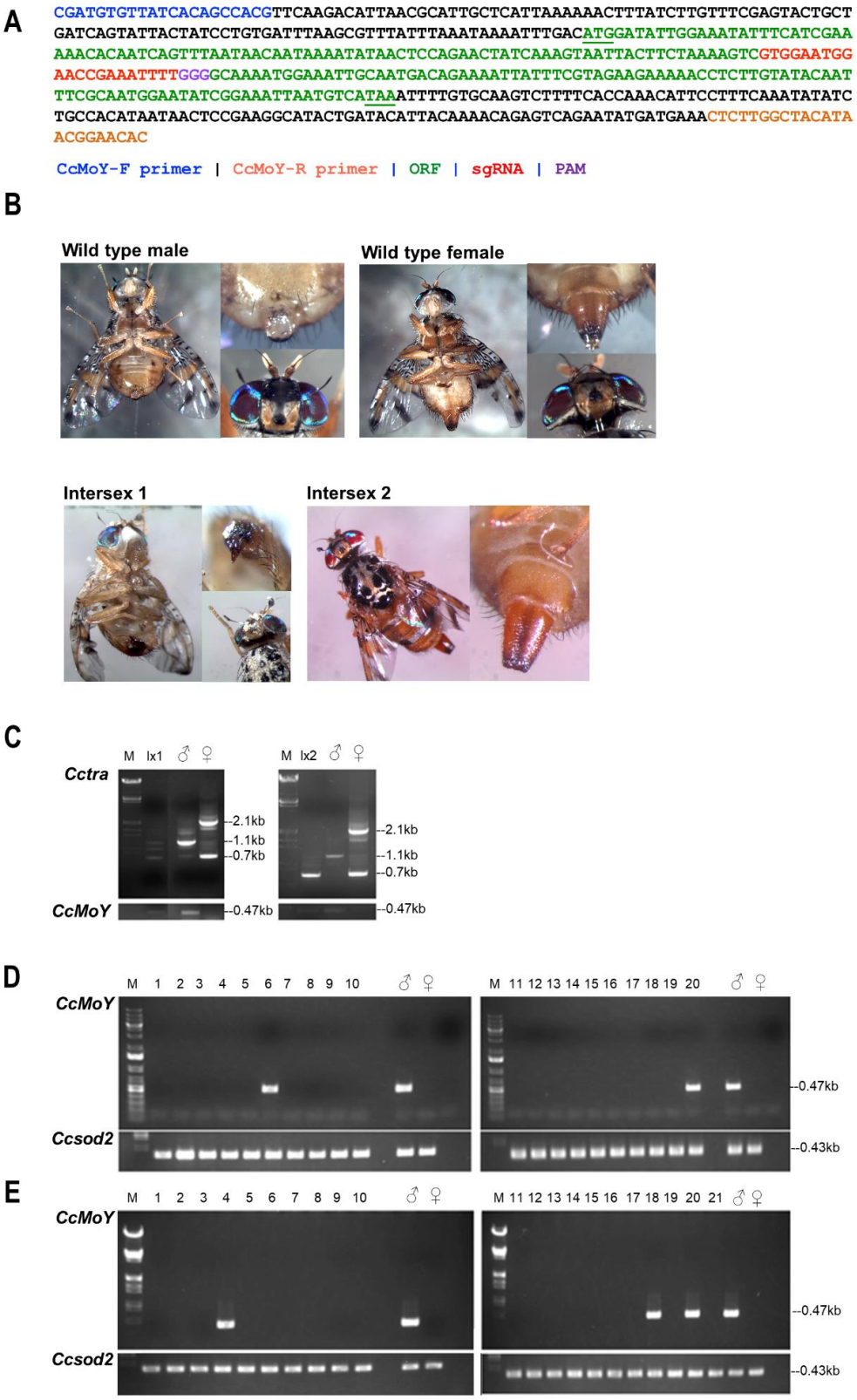

Fig. S4. CRISPR-induced feminization of XY individuals: (A) *MoY* coding sequence (in green) showing the sgRNA target site (in red) and primers *CcMoY*-F and -R. ATG and STOP codons are

underlined. (**B**) 2 adult XY intersexes developed from embryos injected with CRISPR-Cas9 ribonucleoproteins targeting *MoY*: Ix1 (for intersex 1) displays 2 male-specific orbital bristles on the head and malformed ovipositor; Ix2 (for intersex 2) shows 1 male-specific orbital bristle (mosaic) and female ovipositor. Wild type flies are shown in Fig. 2D. (**C**) RT-PCR of *Cctra* (top panels) and of *MoY* (bottom panels) for XX/XY karyotyping: both Ix1 and Ix2 are XY and express a mix of both male and female *Cctra* transcript isoforms. (**D**) PCR on gDNA for molecular karyotyping of the 20 G_0 female flies from the Cas9-sgRNA embryonic injections. *MoY* was used as Y-specific marker (CcMoY-F and CcMoY-R; top panel). *Ccsod2* PCR on genomic DNA was used as positive control (bottom panels). Two G_0 females (numbers 6 and 20) were XY. (**E**) PCR on genomic DNA with primers CcMoY-F and CcMoY-R (top panels) was used for molecular karyotyping of the 21 G_1 females born from the G_0 female #6 crossed with XX males (table S5, fig S5). Three females (numbers 4, 18 and 20) were positive for *MoY*, confirming that these inherited maternally the Y chromosome harboring null mutant of the *MoY* gene (mutant *MoY* DNA sequences were detected in those XY females molecularly; fig. S5A). *Ccsod2* PCR on genomic DNA was used as positive control.

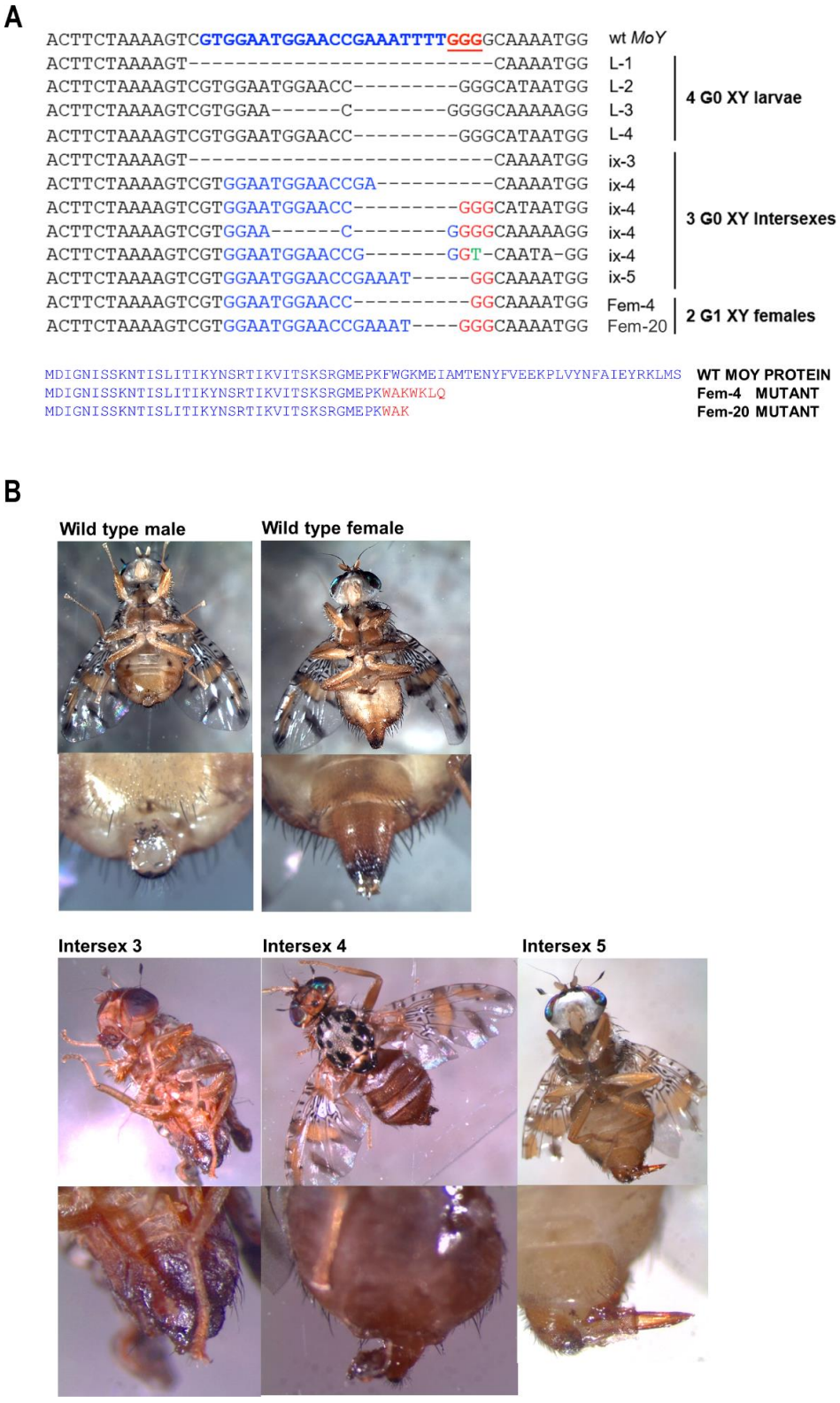

Fig. S5. CRISPR-Cas9–induced disruption of *MoY* causes feminization of XY individuals. (**A**) Injection of Cas9 ribonucleoprotein targeting the putative coding region of *MoY* (see Fig. 1d;

tables S2) induced indels in the proximity of the PAM site, as detected in DNA sequencing from a pool of G_0 larvae, from 3 adult G_0 intersexes (shown in B) and 2 G_1 XY females (Fig. S4E, lanes 4 and 18). The third G_1 XY female was not analyzed for *MoY* sequence because of subsequent DNA degradation of the sample (Fig. S4E, lane 20). The conceptual translation of the 2 mutant *MoY* alleles in G_1 XY fem-4 and fem-20 resulted in truncated MOY proteins respectively 44 and 30 amino acids long. (**B**) Cas9-induced intersexual phenotypes in three other intersexes showed the presence of male-specific orbital bristles on the head (upper row) and deformed genitals similar to ovipositors (lower row). Wild type flies are shown in Fig. 2D.

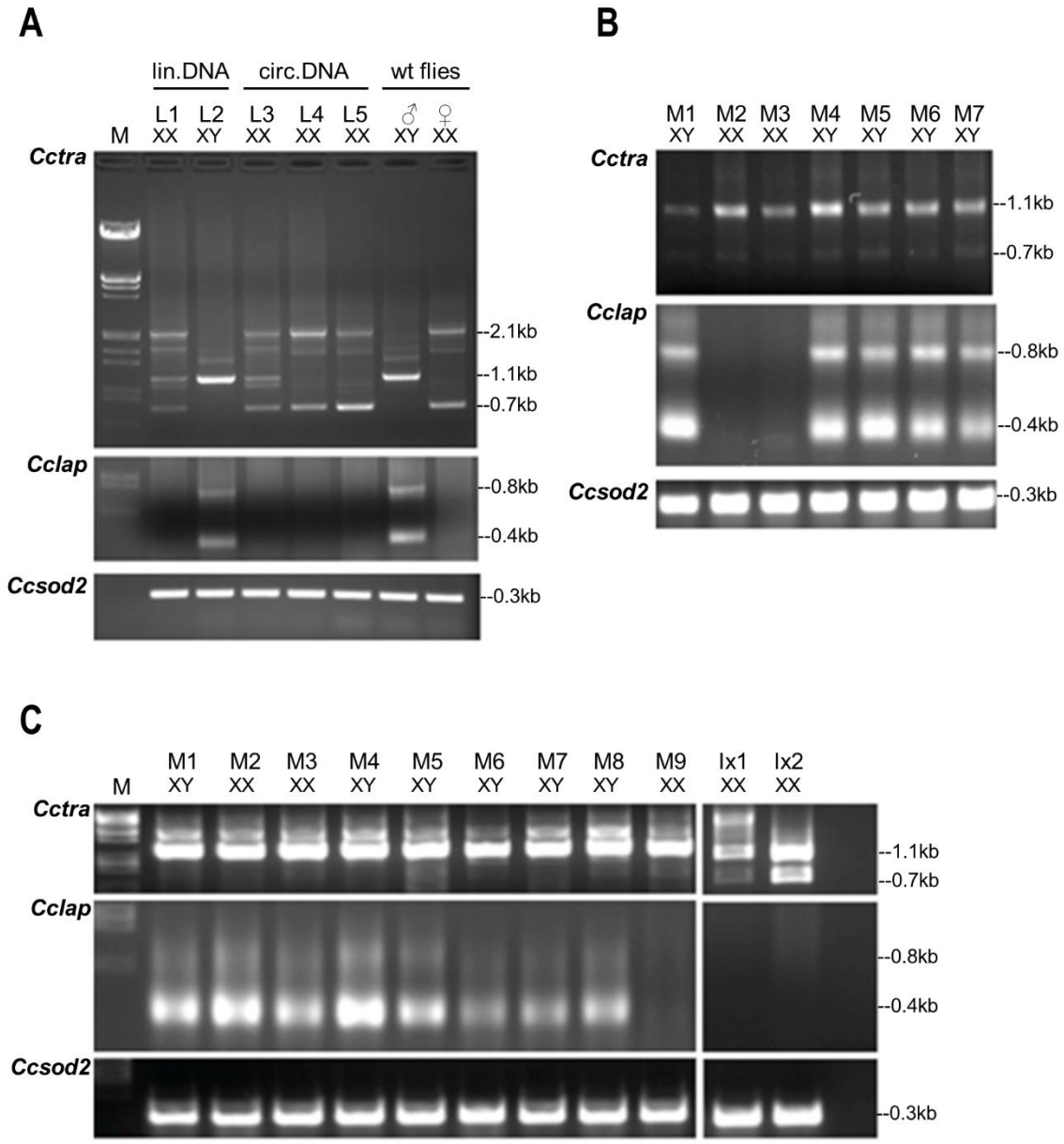

Fig. S6. Masculinization of G_0 individuals following embryonic injection with *MoY* genomic DNA: (**A**) RT-PCR analysis of *Cctra* splicing on larvae developed from embryos injected with linear DNA or plasmid DNA harboring *MoY* (*Cctra* panels). The Y-specific *Cclap* transcript was used to assign karyotype (*Cclap* panels) and *Ccsod2* was used as positive control for cDNA synthesis and control for gDNA contamination (*Ccsod2* panels). L1 and L3 larvae were XX but expressed both male and female isoforms of *Cctra,* in contrast to L4 and L5 XX larvae in which no effect was observed. No effect was observed in L2 XY larva compared to wild type adult XY flies. (**B**) RT-PCR analysis of *Cctra* splicing on 7 phenotypic males developed from embryos injected with linear *MoY* DNA. All XX males (M2 and M3) displayed male specific splicing of *Cctra* similarly to XY males, indicating complete molecular masculinization. (**C**) RT-PCR analysis of *Cctra* splicing on 9 phenotypic males developed from embryos injected with a cloned plasmid containing *MoY* genomic region. The only XX male (M9) displayed male-specific splicing

of *Cctra* similarly to XY males, indicating complete molecular masculinization. Two phenotypic intersexes (Ix1 and Ix2) displayed a mix of male and female *Cctra* splicing isoforms.

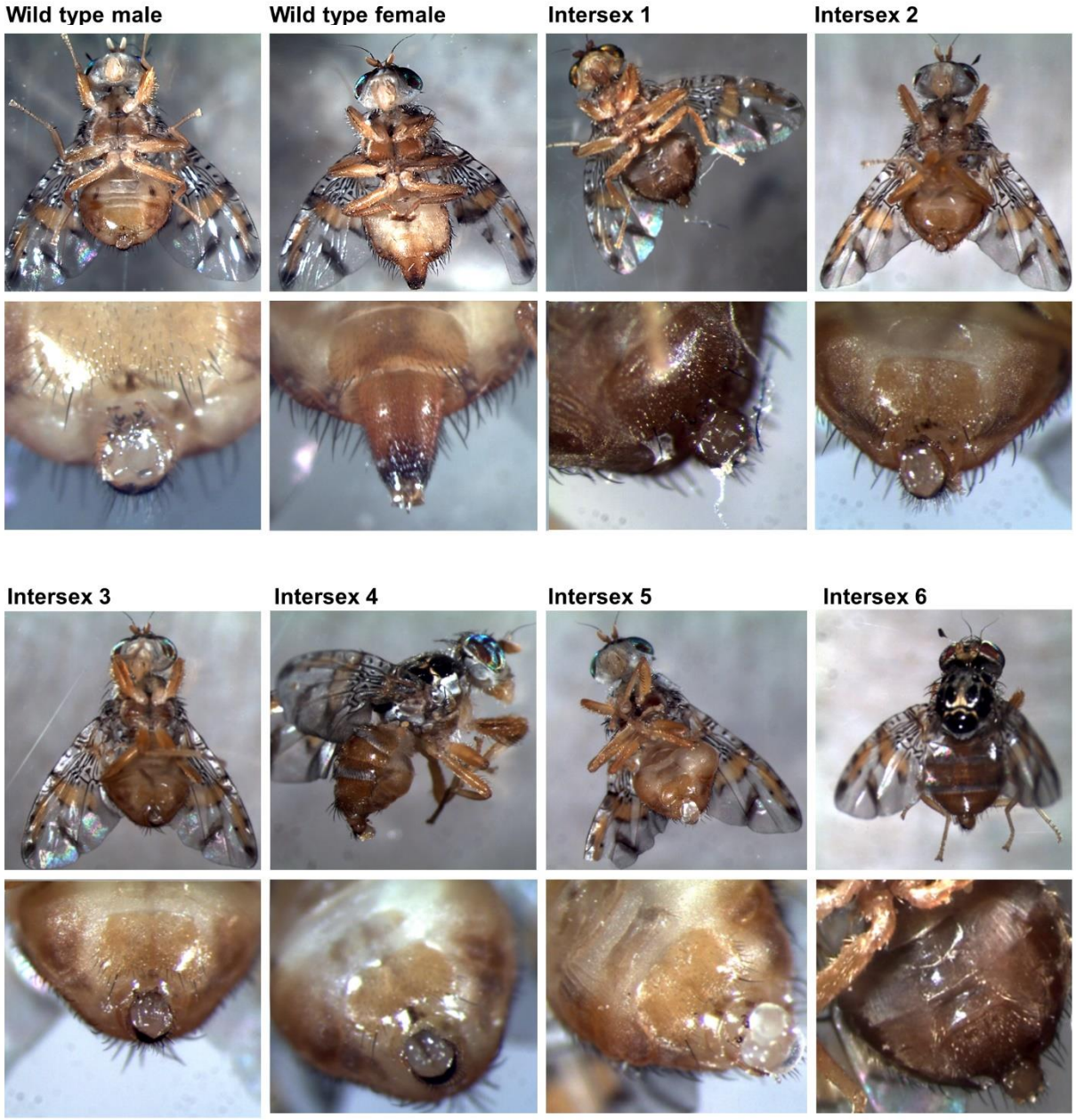

Fig. S7. Embryo injections of linear *MoY* DNA induced adult intersexes. 6 XX intersexes developed from *MoY* DNA injections (see also table S2) showed male external genitalia, but no male-specific head orbital bristles (Ix1, Ix3, Ix4 and Ix5) or only one bristle, indicating sexual mosaicism (Ix2 and Ix6). We concluded that these intersexes are partially masculinized XX individuals. Wild type flies are shown in Fig. 2D.

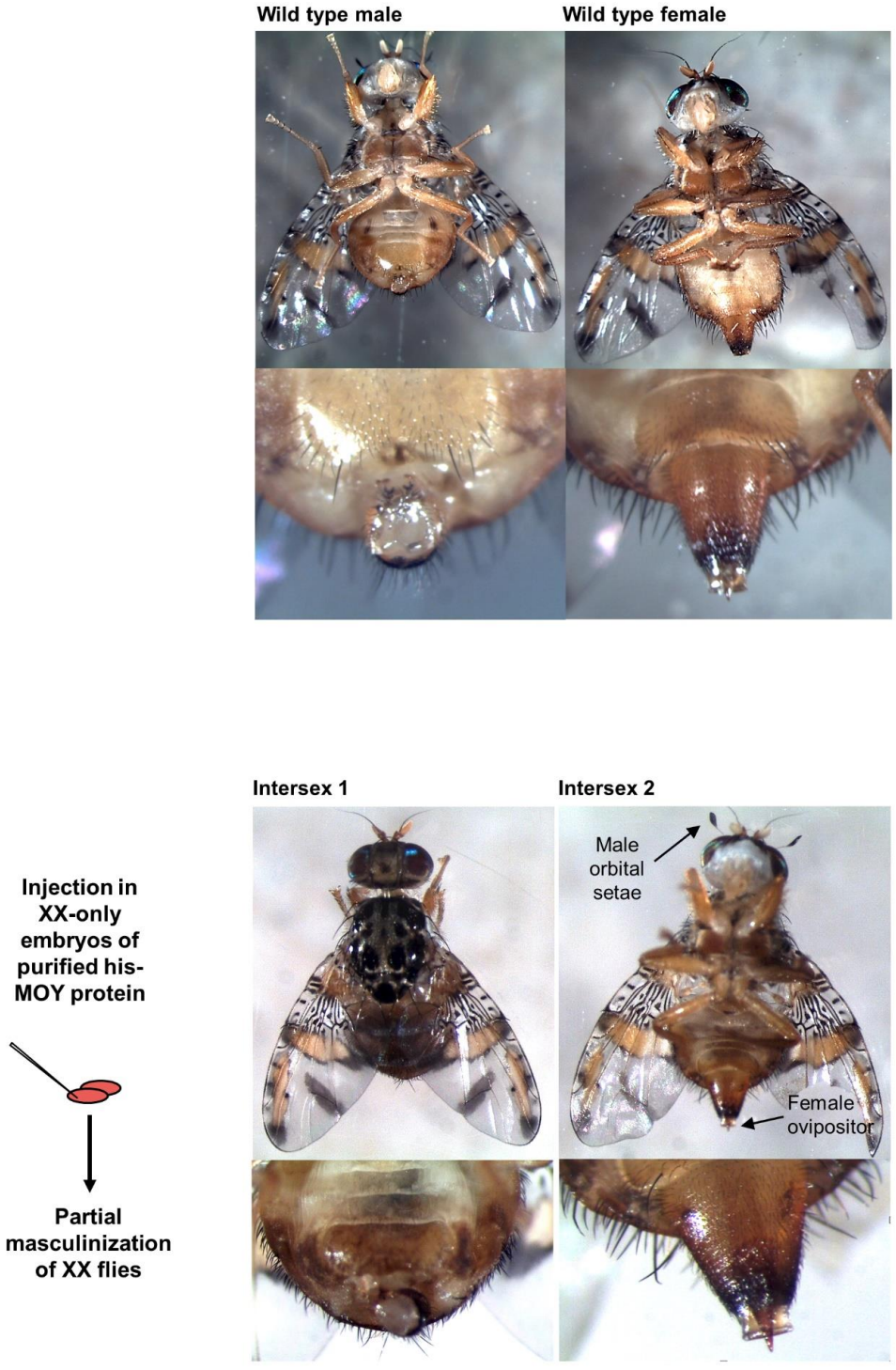

Fig. S8. Embryo injections of MOY protein in XX embryos: purified MOY protein injected into 0-1h AEL embryos (see also table S2) induced partial masculinization of XX flies, either in the posterior region (Ix1), showing male-specific genital terminalia apparatus, or in the anterior region, showing male-specific bristles on the heads. Wild type flies are shown in Fig. 2D.

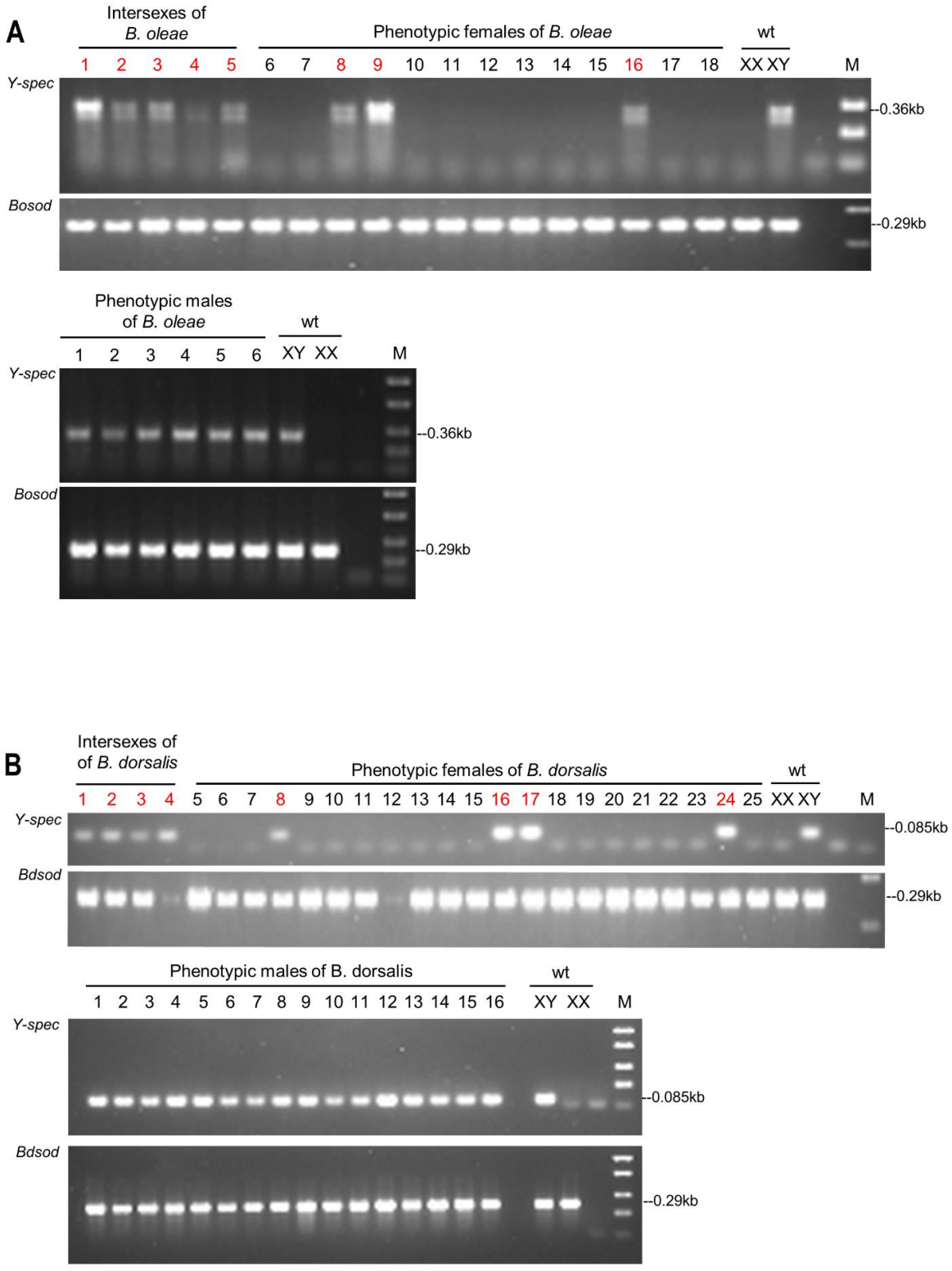

Fig. S9. Molecular karyotyping of intersexes and G0 phenotypic females from *MoY* embryonic RNAi in *B. oleae.* (**A**) **and *B. dorsalis*** (**B**): adult flies (table S2, #7 and 8) were karyotyped by PCR using *MoY* orthologues as a Y-chromosome specific marker. As a positive

control, we used autosomal *sod* (superoxide dismutase) orthologue for each species, respectively. (**A**) In *B. oleae*, 5 intersexes and 3 out of 13 adult females were XY. (**B**) In *B. dorsalis* 4 intersexes and 4 out of 21 adult females were XY. Wild type XX and XY flies are shown as reference for each species. In Lane 12 in B, fly karyotype could not be assigned.

N°	4-8h embryonic Trinity assembly transcript ID	Presence in Medfly Baylor Genome	Presence in Fam18 Canu assembly	Correspondence to candidates from preliminary screening	BLASTx (NCBI) vs *Ceratitis*	BLASTx (NCBI) vs *Drosophila*	Embryonic XY specificity in B. oleae	Male mapped reads	Female mapped reads	Chromosome Quotient
1a	TRINITY_DN40516_c0_g1_i6	No (Shorter highly related sequences)	Multiple		No similarity	No similarity	No similarity	11118	0	0
1b	TRINITY_DN40516_c0_g2_i3	No (Shorter highly related sequences)	Multiple	Lyra	Putative gustatory receptor 59f	No similarity	No similarity	10916	0	0
1c	TRINITY_DN40516_c0_g2_i2	No (Shorter highly related sequences)	Multiple	Lyra	Putative gustatory receptor 59f	Zpg	No similarity	10794	0	0
1d	TRINITY_DN40516_c0_g1_i2	No (Shorter highly related sequences)	Multiple	Lyra	No similarity	No similarity	No similarity	10759	0	0
1e	TRINITY_DN40516_c0_g2_i1	No (Shorter highly related sequences)	Multiple		Putative gustatory receptor 59f	Mabiki	Yes, short and weak	11935	0	0
2	TRINITY_DN38563_c5_g1_i1	No (Shorter highly related sequences)	None		Branchpoint-bridging protein	Quaking related 58E-2 isoform	No similarity	79	0	0
3a	TRINITY_DN40292_c0_g1_i10	No (Shorter highly related sequences)	Multiple		No similarity	No similarity	No similarity	5813	0	0
1f	TRINITY_DN40516_c0_g1_i5	No (Shorter highly related sequences)	None		No similarity	No similarity	No similarity	10249	0	0
3b	TRINITY_DN40292_c0_g3_i1	No	Single	Corvus	No similarity	Lace isoform E	Yes, short	68	0	0
4a	TRINITY_DN40142_c1_g4_i1	No (Highly related paralogous sequences)	Multiple		Histone H2B	Histone H2B	No similarity	872	0	0
5	TRINITY_DN33215_c0_g2_i1	No (Highly related paralogous sequences)	None		Uncharacterized protein	No similarity	Yes, short	3982	0	0
1g	TRINITY_DN40516_c0_g1_i7	No (Shorter highly related sequences)	Multiple		No similarity	No similarity	No similarity	8580	0	0
6	TRINITY_DN40470_c17_g1_i5	No (Highly related paralogous sequences)	None		Cytosol aminopeptidase	Sperm-Leucylaminopeptidase 5	Yes	3182	0	0
4b	TRINITY_DN40142_c1_g1_i5	No (Highly related paralogous sequences)	Multiple		No similarity	No similarity	Yes	274	0	0
4c	TRINITY_DN40142_c1_g4_i2	No (Shorter highly related sequences)	Multiple		Histone H2B	Histone H2B	No similarity	336	0	0
7	TRINITY_DN38104_c3_g2_i2	Yes	None		No similarity	No similarity	No similarity	6253	0	0
8	TRINITY_DN40402_c3_g5_i1	No (Highly related paralogous sequences)	Multiple		No similarity	No similarity	No similarity	16013	0	0
9	TRINITY_DN36540_c0_g1_i1	Yes	None		No similarity	No similarity	No similarity	11903	0	0
10	TRINITY_DN37671_c9_g1_i1	Yes	None		F-Box/SPRY domain-containing protein	No similarity	No similarity	11015	0	0

Table S1. Summary of bioinformatic results for the 19 predicted transcripts: BLASTn analyses showed that most of the 19 transcripts are missing in the currently available Medfly Baylor genome assembly. Only 3 transcripts have hit quality hits in this assembly. In contrast, BLASTn analyses on the male Medfly Canu *Fam18* genome showed that 12 transcripts have corresponding hits, with most present in different contigs suggesting multiple copies. The presence of corresponding transcripts only in the mixed XX/XY but not in the XX-only transcriptome further supported male-specific expression for 15 of them. 4 out of 19 transcripts corresponded to 2 previously selected putative Y-linked male-specific genes (*lyra* and *corvus:* see supplementary text S2). BLASTx analysis on protein databases of *C. capitata* and *D. melanogaster* showed some similarity mostly to short stretch of peptidases, transcription factors, receptors and histone proteins. The number of *Fam18* male and female mapped reads and the chromosome quotient value (CQ) are reported.

Injection Mix #		Injection Mix Type	Karyotypes of injected embryos	Injected embryos	Adults	Unaffected		Feminization of XY		Masculinization of XX	
						XY males	XX females	XY females	XY intersexes	XX males	XX intersexes
Loss of function	#1	dsRNA *MoY*	XX/XY	1217	96	16*	59	14	7	0	0
	#2	dsRNA *MoY*	XX, *wp/wp*; XY-*wp*+, *wp*	260	10	2* (XY-*wp*$^+$)	7	1 (XY-*wp*$^+$)	0	0	0
	#3	CRISPR/Cas9 vs *MoY*	XX/XY	250	32	7*	18	2	5	0	0
Gain of function	#4	Linear 5Kb *MoY* DNA	XX/XY	310	28	16	3°	0	0	3	6
	#5	*MoY* 5Kb plasmid	XX/XY	190	16	8	5°	0	0	1	2
	#6	His-MOY protein	XX	428	31	-	25	-	-	0	6
Loss of function	#7	dsRNA *BoMoY*	XX/XY	550	24	6	10	3	5	0	0
	#8	dsRNA *BdMoY*	XX/XY	540	41	16	17	4	4	0	0

Table S2. Summary of microinjections into *C. capitata* (#1-6), *B. oleae* (#7) and *B. dorsalis* embryos (#8): in red are indicated numbers of flies showing partial (intersex) or full sexual transformations. In injection set 2#, a Y-marked brown pupae strain (*Vienna8*), carrying a white pupae (*wp*) recessive mutation on an autosome was used and an XY female eclosed from a brown pupa. The male flies marked with * in strongly female-biased progenies (#1, #2 and #3) were assigned to XY karyotype without molecular analyses and are considered escapers. Similarly, the female flies marked with ° in male-biased progenies (#4 and #5) were assigned to XX karyotype without molecular analyses. One XY female from injection set#1 when crossed with 3 XX males gave rise to a G_1 progeny of 1 male and 9 females. When crossed with 3 XX males, the single XY female (born from a brown pupae) of injection set #2 gave rise to a G_1 progeny of 1 male and 1 female. One XY female of injection set #3 when crossed with 3 XX gave rise to a progeny of 3 XY females and 18 XX females. The *Bactrocera* female, male and intersexual progenies from injection sets #7 and #8 were molecularly karyotyped, except for one individual in set #8 (fig. S9B).

Injection Mix #	Injection Type #	Injected embryo karyotype	# of injected embryos	# of adults	# of males	# of females	# of intersexes	wt XX females	wt XY males	XY females	XY intersexes
#1a	dsRNA targeting *MoY*	XX/XY	517	44	11	28	5	25	11*	3	5
#1b	dsRNA targeting *MoY*	XX/XY	200	19	1	18	0	14	1*	4	-
#1c	dsRNA targeting *MoY*	XX/XY	500	33	4	27	2	20	4*	7	2
total	total	XX/XY	1217	96	16	73	7	59	16	14	7
buffer only	buffer only	XX/XY	110	50	25	25	-	-	-	-	-

Table S3. Detailed numbers of three embryonic *MoY* eRNAi microinjection sets: the *MoY*-eRNAi injection experiment #1 reported in table S2 is the sum of 3 independent injection experiments reported here. In red are indicated the number of flies showing partial (intersex) or full sexual transformation. From eRNAi injection experiment #1a: targeting *MoY*, the G0 progeny was composed of 25 XX females and 11 XY males, as well as 3 XY females and 5 XY intersexes. The ratio of 25 XX versus 19 XY flies is not significantly different with respect to a 1:1 ratio (chi-squared test). From eRNAi injection experiment #1b: targeting *MoY*, the ratio of 14 XX versus 5 XY flies is significantly deviant with respect to 1:1 ratio (chi-squared test). In contrast, eRNAi experiment #1c showed a ratio of XX versus XY flies close to the expected chi-square value. The male flies marked with *, among female-biased progeny, were assigned to XY karyotype, considering them as *MoY* RNAi escapers.

Female #	Karyotype	Pupae	Adult flies	Males	Females
26	XY	11	10	1	9
29	XX	17	10	0	10
30	XX	0	0	0	0
31	XX	45	44	0	44
32	XX	28	24	0	24
33	XX	26	17	0	17
34	XX	21	14	0	14
35	XX	56	54	0	54
36	XX	6	4	0	4
37	XY	0	0	0	0
38	XY	1	0	0	0
39	XY	0	0	0	0
40	XX	0	0	0	0
41	XX	14	11	0	11
42	XX	0	0	0	0
43	XX	23	14	0	14
44	XY	0	0	0	0
45	XX	0	0	0	0
46	XX	0	0	0	0
47	XX	0	0	0	0
48	XX	0	0	0	0
49	XX	0	0	0	0
50	XX	3	3	0	3

Female #	Karyotype	Pupae	Adult flies	Males	Females
51	XX	0	0	0	0
52	XX	2	2	0	2
53	XY	0	0	0	0
54	XX	17	15	0	15
55	XX	0	0	0	0
56	XX	34	32	0	32
57	XX	0	0	0	0
58	XX	0	0	0	0
59	XX	8	8	0	8
60	XY	0	0	0	0
61	XX	0	0	0	0
62	XX	0	0	0	0
63	XX	35	32	0	32
64	XY	0	0	0	0
65	XX	0	0	0	0
66	XX	0	0	0	0
67	XY	0	0	0	0
68	XX	32	26	0	26
69	XY	0	0	0	0
70	XY	0	0	0	0
71	XX	3	0	0	0
72	XX	16	14	0	14
73	XX	0	0	0	0

Table S4. Fertility test of G_0 XX and XY females from *MoY*-eRNAi: results from crosses of 46 out of 73 G_0 females to XX males, derived from *MoY*-eRNAi. The females are numbered referring to the PCR karyotyping reported in fig. S2 and females that are XY are highlighted in blue rows. One XY female (# 26) gave adult progeny, including 1 male and 9 females. One XY female (# 38) gave 1 pupa which failed to develop into an adult fly

.

Female #	Karyotype	G1 Pupae	G1 adults	Male G1	Female G1
1	XX	70	19	0	19
2	XX	0	0	0	-
3	XX	56	53	0	53
4	XX	31	23	0	23
5	XX	31	23	0	23
6	XY	25	21	0	21
7	XX	0	0	0	0
8	XX	10	8	0	8
9	XX	0	0	0	0
10	XX	2	0	0	0
11	XX	0	0	0	0
12	XX	63	50	0	50
13	XX	0	0	0	0
14	XX	95	89	0	89
15	XX	0	0	0	0
16	XX	37	30	0	30
17	XX	43	42	0	42
18	XX	0	0	0	0
19	XX	28	27	0	27
20	XY	0	0	0	0

Table S5. *MoY-Cas9* mutant G_0 XY females can also be fertile, as XY females from eRNA-*MoY* (table S4). 20 adult G_0 females, derived from *MoY*-Cas9 micro-injections, were individually crossed to 3 XX males. PCR karyotyping revealed that 2 out of 20 females were XY (#6 and #20 shaded in blue rows). One of the 2 XY females (# 6) had G_1 offspring (21 individuals, 3 XY and 18 XX; Fig. S4E) which as expected were female only. The second G_0 XY female (# 20) died a few days after hatching. The reduced fertility of XY females, produced by *MoY*-eRNAi (table S4) or *MoY*-Cas9 targeting (this table), is possibly due to cellular mosaicism of perturbed sex determination. The bias in favor of XX offspring could be due either to reduced viability of Y-carrying eggs or distorted segregation of X and Y chromosomes during female meiosis.

ID	length	*Fam18* male_counts	*Fam18* female_counts	ISPRA male_counts	ISPRA female_counts
Cctra-2 (single copy, autosomal gene)	1113	31	25	40	48
pY114 (Y-specific repetititve element)	1405	160	0	232	0
pm11 (Y-specific repetititve element)	2733	148***	0	317***	0
5Kb (Y-specific repetititve element)	5642	0***	0	2344***	6
pM21 (Y-specific repetititve element)	1380	189***	0	327***	0
Total reads		242122621	246407224	322938668	275803598

Table S6. Summary of mapping analysis using the *Fam18* and *ISPRA* strains: the number of mapping reads is shown comparing four Y-chromosome specific repetitive elements and *Cctra*, as a control, from WGS reads of males and females of the *Fam18* and *ISPRA* strains. *** $p < 0.001$, Fisher Exact Test comparing *Fam18* male read counts versus *ISPRA* male read counts.

Total sequenced reads	179,917,948
Trimmomatic filtered reads	168,144,620
Ribosomal and mitochondrial depleted reads	148,886,218
Total assembled bases (bp)	187,376,939
Trinity assembled transcripts	213,154
GC content (%)	37.25
Median transcript length (bp)	399
Average transcript length (bp) 1253.79	879
Transcript N50 (bp)	1,828
Shortest transcript length (bp)	201
Longest transcript length (bp)	17,588
Trinity transcripts > 1Kb	48,375
Trinity transcripts > 2Kb	24,380

Table S7. Summary of sequencing and transcriptome assembly statistics: Illumina short read sequencing and assembly statistics of the Medfly transcript catalogue produced using 4-8h AEL embryonic RNA and the Trinity *de novo* assembler.

Primer Name	Sequence (5' – 3')
Bo_MoY_F1	GGAAGTAGCTAACTAAATGATTG
Bo_MoY_R1	ATTATTCGTCGTTCAGAAGTCG
Bd_MoY_F1	AAATGATATAGAAGAGCATGG
Bd_MoY_R3	TCCGCTCCGGAAAAATTTAGG
Bt_Bj_MoY_F1	CATTGAGACTTCTAGTAATTT
Bt_Bj_MoY_R1	GCTCACGAAGAATCAATGC
T7_CcMoY-F	TAATACGACTCACTATAGGGAGACGATGTGTTATCACAGCCACG
T7_CcMoY-R	TAATACGACTCACTATAGGGAGAGTGTTCCGTTATGTAGCCAAGAG
T7_BdMoY-F	TAATACGACTCACTATAGGGAGAAAATGATATAGAAGAGCATGGG
T7_BdMoY-R	TAATACGACTCACTATAGGGAGATCCGCTCCGGAAAAATTTAGG
T7_BoMoY-F	TAATACGACTCACTATAGGGAGAGGAAGTAGCTAACTAAATGATTG
T7_BoMoY-R	TAATACGACTCACTATAGGGAGAATTATTCGTCGTTCAGAAGTCG
CRISPR-CcMoY-F	GAAATTAATACGACTCACTATAGGGAATGGAACCGAAATTTTGGTTTTAGA GCTAGAAATAGC
CRISPR-Reverse	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTA TTTTAACTTGCTATTTCTAGCTCTAAAAC
Cas9-MoY F	CGATGTGTTATCACAGCCACG
Cas9-MoY R	GTGTTCCGTTATGTAGCCAAGAG
CcMoY A+	ACGGAACACATGCTAGCAGA
CcMoY A-	TTGCCCCAAAATTTCGGTTCC
CcYF	GCTCGAAGACATGCATTGAA
CcYR	GACGGTAAGTGCCATTCGTT
Cctra 164+	CAGTGGTTCGGTTCGGAAG
Cctra 900-	GGACAATATCGACATCATGGA
Cclap+	AAGGACTTGTGATTGGATTG
Cclap-	ATGCCGTCGTCCAACATC
Ccsod2+	TGCTCCGAGAACGTTCACG
Ccsod2-	TCATCGGTCAATTCGTGCAC
BdMoY-F	AAATGATATAGAAGAGCATGGG
BdMoY-R	CGTTGAGAAGAGGTTAGTAT
BoMoY-F	GGAAGTAGCTAACTAAATGATTG
BoMoY-R	ATTATTCGTCGTTCAGAAGTCG
BdSOD-F	CCTTAATAGTCACACTAGCC
BdSOD-R	CACATGGAAACCGTGTTTAC
BoSOD-F	CCTTAATAGTCACATTGGCC
BoSOD-R	CACATGGAAACCGTGCTTAC
Bd-tra_Fwd	GAAGTTGTTATTAAGCGTAGATTCGG
Bd-tra_Rev	CTTTCCCGTTCGCGTTTACTATTG
Bo-tra_Fwd	TACCGCACAAATTCAATCCATG
Bo-tra_Rev	GAGCATCTGCGGTATTTGTAAC
New_MoY_F1	TATTTCATCGCTCGCGAAATTGAG
New_MoY_R	CAAAATTACAAACGTTTGGCATTTG
Gen_CcMoY_F	GTCCTCCTGTTTGTCACTTAC
Gen_CcMoY_R	GGAGGCAAGTGAATCTATTTGG
CcOrf2fw (NcoI)	CATGCCATGGATATTGGAAATATTTCATCG
CcOrf2rv(XhoI)	GATGCTCGAGTGACATTAATTTCCGATATTCTATT
atub+	CGCATTCATGGTTGATAACG
atub-	GGGCACCAAGTTAGTCTGGA

Table S8. Primer sequences.

Gene name	Primer pair (5'-3')	Amplicon Size in bp	Efficiency	R^2
β-TUB (TBB1)	F=TCTCTACCAGTTGATGCAC R=CCGACAGAATAATGAACAC	105	96.30%	0.993
CcMoY	F=GTCTGTTCACCAAACATTCCTTTC R=TGTGTTCCGTTATGTAGCCAAG	111	104.70%	0.996
RPL19	F=AACAAACGTGTACTGATGG R=CACGTACTTTATGTCGTCTG	103	90.60%	0.999

Table S9. Primer pairs used for qRT-PCRs.

Data S1. (separate file)
qRT-PCR expression analysis data.

Data S2. (separate file)
Transcript quantifications of 4-8h Medfly RNA-seq data (against embryonic transcriptome).

Data S3. (separate file)
EdgeR differential expression analysis output table of 4-8h Medfly RNA-seq data.

Data S4. (separate file)
Bactrocera spp. transcriptome assembly statistics.

REFERENCES

1. Y. Qin, D. R. Paini, C. Wang, Y. Fang, Z. Li, Global establishment risk of economically important fruit fly species (Tephritidae). *PLoS One* **10**, e0116424 (2015).
2. J. H. V. A. Dyck, A. S. Robinson, Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management. (2005).
3. D. Bopp, G. Saccone, M. Beye, Sex determination in insects: variations on a common theme. *Sex Dev* **8**, 20-28 (2014).
4. D. Bachtrog *et al.*, Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, (2014).
5. A. Pane, M. Salvemini, P. Delli Bovi, C. Polito, G. Saccone, The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* **129**, 3715-3725 (2002).
6. M. Hediger *et al.*, Molecular characterization of the key switch *F* provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. *Genetics* **184**, 155-170 (2010).
7. A. Papanicolaou *et al.*, The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol* **17**, 192 (2016).
8. U. Willhoeft, G. Franz, Identification of the sex-determining region of the *Ceratitis capitata* Y chromosome by deletion mapping. *Genetics* **144**, 737-745 (1996).
9. P. Gabrieli *et al.*, Sex and the single embryo: early development in the Mediterranean fruit fly, *Ceratitis capitata*. *BMC Dev Biol* **10**, 12 (2010).
10. A. B. Hall *et al.*, Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics* **14**, 273 (2013).
11. A. Meccariello *et al.*, Highly efficient DNA-free gene disruption in the agricultural pest *Ceratitis capitata* by CRISPR-Cas9 ribonucleoprotein complexes. *Sci Rep* **7**, 10061 (2017).
12. M. N. Krosch *et al.*, A molecular phylogeny for the Tribe Dacini (Diptera: Tephritidae): systematic and biogeographic implications. *Mol Phylogenet Evol* **64**, 513-523 (2012).

13. A. B. Hall *et al.*, SEX DETERMINATION. A male-determining factor in the mosquito *Aedes aegypti*. *Science* **348**, 1268-1270 (2015).
14. A. Sharma *et al.*, Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene CWC22. *Science* **356**, 642-645 (2017).
15. E. Krzywinska, N. J. Dennison, G. J. Lycett, J. Krzywinski, A *maleness* gene in the malaria mosquito *Anopheles gambiae*. *Science* **353**, 67-69 (2016).
16. F. Criscione, Y. Qi, Z. Tu, *GUY1* confers complete female lethality and is a strong candidate for a male-determining factor in *Anopheles stephensi*. *Elife* **5**, (2016).
17. N. P. Kandul *et al.*, Transforming insect population control with precision guided sterile males with demonstration in flies. *Nat Commun* **10**, 84 (2019).
18. V. M. Gantz, E. Bier, Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **348**, 442-444 (2015).
19. K. Kyrou *et al.*, A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat Biotechnol*, (2018).
20. P. A. Papathanos, N. Windbichler, Redkmer: An Assembly-free pipeline for the identification of abundant and specific X-Chromosome target sequences for X-Shredding by CRISPR endonucleases. *CRISPR J* **1**, 88-98 (2018).
21. J. E. Anleitner, D. S. Haymer, Y enriched and Y specific DNA sequences from the genome of the Mediterranean fruit fly, *Ceratitis capitata*. *Chromosoma* **101**, 271-278 (1992).
22. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
23. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
24. M. Salvemini *et al.*, *Ceratitis capitata transformer-2* gene is required to establish and maintain the autoregulation of *Cctra*, the master gene for female sex determination. *Int J Dev Biol* **53**, 109-120 (2009).
25. M. Salvemini *et al.*, *De novo* assembly and transcriptome analysis of the Mediterranean fruit fly *Ceratitis capitata* early embryos. *PLoS One* **9**, e114191 (2014).
26. B. J. Haas *et al.*, *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512 (2013).

27. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
28. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).
29. D. S. Holmes, J. Bonner, Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. *Biochemistry* **12**, 2330-2338 (1973).
30. H. Li, BFC: correcting Illumina sequencing errors. *Bioinformatics* **31**, 2885-2887 (2015).
31. J. A. Chapman *et al.*, Meraculous: *De novo* genome assembly with short paired-end reads. *PLoS One* **6**, e23501 (2011).
32. T. Hackl, R. Hedrich, J. Schultz, F. Forster, proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**, 3004-3011 (2014).
33. V. Petrella *et al.*, *De novo* assembly and sex-specific transcriptome profiling in the sand fly *Phlebotomus perniciosus* (Diptera, Phlebotominae), a major Old World vector of Leishmania infantum. *BMC Genomics* **16**, 847 (2015).
34. A. Kato, P. S. Albert, J. M. Vega, J. A. Birchler, Sensitive fluorescence in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotech Histochem* **81**, 71-78 (2006).
35. L. Z. Carabajal Paladino, P. Nguyen, J. Sichova, F. Marec, Mapping of single-copy genes by TSA-FISH in the codling moth, Cydia pomonella. *BMC Genet* **15 Suppl 2**, S15 (2014).
36. V. Krylov, T. Tlapakova, J. Macha, Localization of the single copy gene *Mdh2* on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet Genome Res* **116**, 110-112 (2007).
37. E. Sagri *et al.*, Housekeeping in Tephritid insects: the best gene choice for expression analyses in the medfly and the olive fly. *Sci Rep* **7**, 45634 (2017).
38. T. G. Montague, J. M. Cruz, J. A. Gagnon, G. M. Church, E. Valen, CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-407 (2014).
39. M. Salvemini, U. Mauro, S. Velaeti, C. Polito, G. Saccone, A new *Minos* vector for eye-specific expression of *white*+ marker in *Ceratitis capitata* and in distantly related dipteran species. *Insect Mol Biol* **15**, 341-349 (2006).
40. G. K. Banks *et al.*, A second *superoxide dismutase* gene in the Medfly, *Ceratitis capitata. Genetics* **140**, 697-702 (1995).